Author:
Pemberton, Jonathan

Title:
Untargeted Chemometric Characterisation of High Resolution Orbitrap Mass Spectra of Riverine and Point Source Dissolved Organic Matter

Untargeted Chemometric Characterisation of High Resolution Orbitrap Mass Spectra of Riverine and Point Source Dissolved Organic Matter

by

Jonathan Andrew Pemberton

A dissertation submitted to the University of Bristol, U.K in accordance with the requirements for award the degree of Doctor of Philosophy in the faculty of Science

School of chemistry

November 2018

Words 53,980

Abstract

Dissolved organic matter (DOM) is found in all natural aquatic systems and consists of a complex mixture of compounds that play a vital role in the ecosystem. However, anthropogenic compounds discharged into the water cycle threaten future water security and the aquatic ecosystem. These micropollutants are often low in concentration but their chronic exposure to organisms has been shown to have a wide range of adverse effects. This has led to a number of high profile media articles which demonstrate the growing public concern and awareness of the impact that micropollutants are having on the environment. Therefore, the comprehensive identification of known and unknown micropollutants and the characterisation of their sources is crucial to understanding the impact that these compounds may have on the environment. The chemical complexity of the DOM has until recently precluded detailed investigation of its chemistry at the molecular level. Chemical investigations of organic compounds have largely revolved around targeted analyses of priority pollutants for regulatory purposes. Comprehensive or untargeted analyses of DOM chemistry has until recently been an unattainable goal. However, recent advances in mass spectrometry, particularly routine liquid introduction using electrospray ionisation combined with high resolution mass spectrometry, have introduced the possibility of investigating DOM chemistry in hitherto unattainable detail.

This thesis aimed to develop and apply such techniques, specifically DI-Orbitrap HRMS and HPLC-Orbitrap HRMS, in an untargeted approach to investigate the DOM composition of point sources in comparison to the background river DOM at three UK sewage treatment works. A critical aspect of the method development was the implementation of a novel combination of statistical tools required to interpret the complex distribution of ions revealed in both the background river and sewage works effluent DOM.

DIMS revealed differences and similarities in the point source and background river DOM at all three sites. The complexity of the DI-HRMS spectra made a direct manual comparison between spectra, containing many thousands of individual ions, impractical. However, by using multivariate statistics including PCA and hierarchical cluster analysis, compositional differences between different DOM extracts were clearly revealed. Heatmaps were shown to provide an effective visualisation method to compare individual ions across different complex DI-HRMS spectra. Kruskal-Wallis analysis provided an essential data reduction step to determine the discriminating ions between different DOM spectra. The identification of compounds giving rise to these ions was achieved in some cases using HPLC-MS and HPLC-MS/MS. This approach highlighted compounds, which had been previously identified in sewage effluent demonstrating that this new approach prioritises a relevant list of compounds. Significantly, the method revealed a number of pollutants, not previously identified, as originating from a sewage treatment works. In particular, the combined use of DI-HRMS prior to HPLC-MS allowed the identification of oligomeric molecular species, which were not apparent using HPLC-MS alone.

Overall, the research presented in this thesis demonstrates a novel methodology of potentially wide utility for comprehensively assessing the chemistry of DOM from point sources in order to identify both known and previously unknown micropollutant contributions to aquatic environments.

Acknowledgements

Thank you to both my supervisors Prof. Richard Evershed and Prof. Penny Johnes. Thank you, Richard for your advice and support throughout my PhD project and challenging me to push my project further. I never thought a chance meeting would lead me to doing a summer project and PhD. Thank you, Penny, for your unending enthusiasm and ambition.

My PhD was funded as a NERC CASE studentship between the University of Bristol and the Food and Environmental Research Agency (FERA). I would like to thank FERA for the opportunity to network at your annual student conference and the opportunity to present at the UK Water Mutual Aid Meeting in 2016. In particular, my industrial supervisor Dr Adrian Charleton and Mike Dickinson.

I would like to thank the DOMAINE project chemistry team: Dr Charlotte Lloyd, Catherine McIntyre, Michaela Reay and Leo Rivera for their support in: impromptu sampling trips, discussing ideas and all their advice throughout the project. I would also like to thank the DOMAINE geography team: Chris Yates and Alun Owen for taking me out into the field and collecting samples.

Thanks to Dr Chris Arthur for his training on using the Orbitrap and answering calls on weekends for over the phone troubleshooting.

I would like to thank my project student Laurie Brummitt who carried out the passive sampling work as her MSci project and the excellent effort she put in to complete the project.

I would like to thank Prof. Davey Jones for hosting me at Bangor university for three weeks to collect and extract samples from the Conwy. Also, thanks to Dr Helen Glanville and Fran Brailsford for showing me around Wales and hopefully we will collaborate again in the future.

The OGU has been a brilliant place to work over the last 4 years and I would like to thank all members past and present as I will undoubtedly forget people if I try and name you all. You all contribute to making the OGU a supportive and ambitious environment.

I cannot thank my parents Yvonne and Phillip Pemberton enough for their support, hot meals, listening to me talk about rivers. In particular, my dad for proof reading my thesis, helping me rephrase my ideas, taking the time to listen and your encouragement. Finally, I would like to thank Florence for putting up with me, working weekends or just popping to the office to check on some data.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Table of Contents

List of figures

List of tables

Abbreviations

AIDS - Acquired immunodeficiency syndrome

API – Application programming interface

APPI – Atmospheric pressure photoionisation

DEFRA – Department of environment, food and rural affairs

DI – Direct infusion

DI-HRMS – Direct infusion high resolution mass spectrometry

DOC – Dissolved organic carbon

DOM – Dissolved organic matter

EIC – Extracted ion chromatogram

ESI – electrospray ionisation

FERA – Food and environmental research agency

FTICR-MS – Fourier transform ion cyclotron mass spectrometer

GC – Gas chromatography

HESI – Heated electrospray ionisation

HIV – human immunodeficiency virus

HRMS – High resolution mass spectrometry

HPLC – High performance liquid chromatography

LLE – Liquid-liquid extraction

MS – Mass spectrometry

MS/MS – Tandem mass spectrometry

NHS – National health service

NMR – Nuclear magnetic resonance spectroscopy

NDIR – Nondispersive Infrared detector

PCA – Principal Component analysis

PC – Principal component

pe – Population equivalents

QC - Quality control

RO/ED – Reverse osmosis coupled with electrodialysis

SPE – Solid phase extraction

SUVA – Specific ultraviolet absorbance

TIC – Total ion chromatogram

TOC – Total organic carbon

UV-vis – Ultraviolet-visible spectroscopy

Chapter 1

Introduction

Chapter 1. Introduction

1.1 Dissolved organic matter in aquatic ecosystems

Fresh surface water is a fundamental resource, not only for drinking water and irrigation, but also for supporting terrestrial and freshwater aquatic ecosystems. However, only 0.03% of the Earth's water exists as surface freshwater (Shiklomanov, 1993). Dissolved organic matter (DOM) is ubiquitous in all aquatic systems. As precipitation falls on an area of land, the topography and geology of the landscape determine the drainage of that water. As the water moves it aggregates forming a network of streams, which combine as the water travels downstream to form rivers. These rivers will drain into a particular outlet, for example the sea or a lake. The hydrological processes through the landscape will mobilise soluble organic compounds into these streams, rivers and lakes.

The organic carbon present in aquatic ecosystems forms a crucial part of the global carbon cycle for the transportation, processing and storage of carbon as shown in Figure 1.1. Both marine and inland waters play a vital role in the carbon cycle, acting as a sink. Rivers provide the largest transportation network for terrestrial derived organic matter. Approximately 2.7 Pg of carbon is transported and processed in rivers annually (Degens *et al.*, 1991; Battin *et al.*, 2009). However, in a river the typical concentration of DOM ranges from 0.1 mg L^{-1} – 10 mg L^{-1} (Minor *et al.*, 2014). DOM is an integral part of a river's ecosystem, it acts as a natural sun screen and can be used by autotrophs as a nutrient.

Anthropogenic pollution is known to affect the aquatic ecosystem. Increased nutrient loads from agriculture and sewage disposal can cause large algal blooms resulting in anoxia in rivers and lakes (Smith *et al.*, 1998; Novotny, 1999; Pinckney *et al.*, 2001). This depletion in the oxygen supply in the water can cause large scale fish kills causing irreparable damage to the ecosystem (Novotny, 1999; Pinckney *et al.*, 2001; Nikinmaa, 2014). Eutrophication renders the water unusable as a water resource, not only for abstraction for human consumption, but also for terrestrial organisms which inhabit the surrounding landscape.

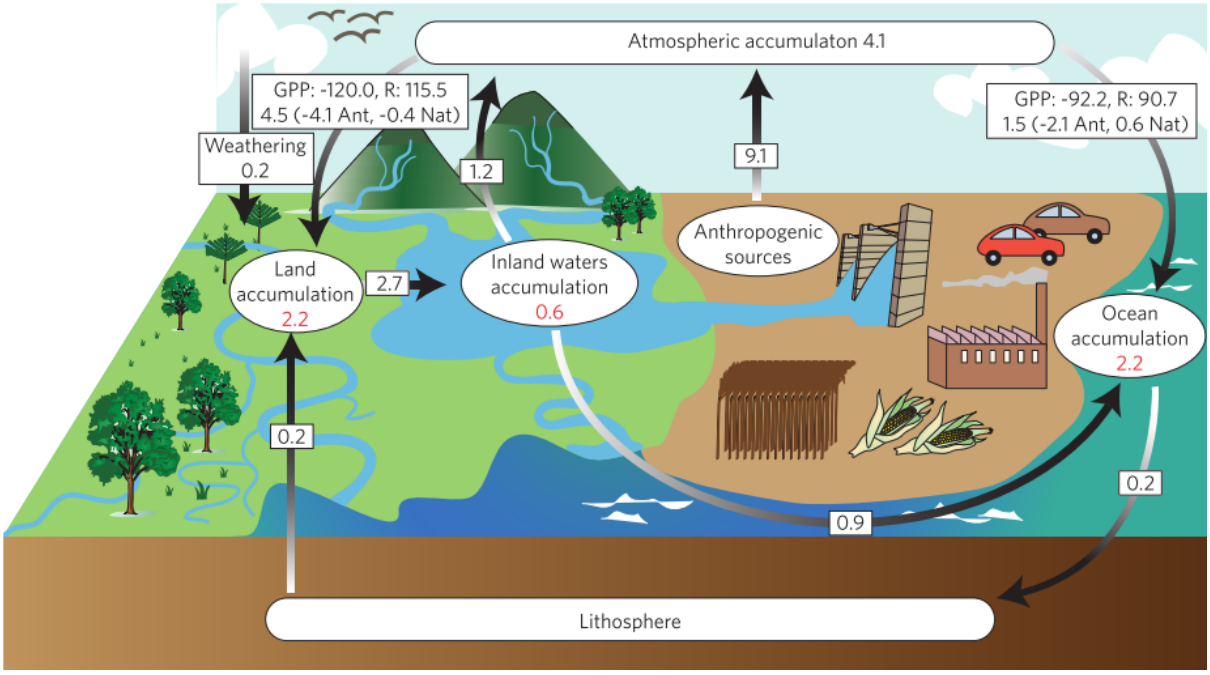


Figure 1.1. The carbon cycle adapted, units are Pg C yr^{-1}. Rates of changes are highlighted in red and fluxes between pools are in black. From Battin *et al.*, (2009).

However, there is a growing body of research showing the adverse effects of individual dissolved micropollutants entering the surface water through human activity (Jobling *et al.*, 1998; Malaj *et al.*, 2014; Capaldo *et al.*, 2018). A micropollutant is defined as a synthetic substance, which are in low concentration in the environment typically less than micrograms per litre. Despite being in low concentration, micropollutants can still have adverse effects on the surrounding ecosystem. These effects are subtler than eutrophication however, the chronic exposure of biologically active micropollutants to aquatic organisms has been shown to have damaging long term effects. These effects can be wide ranging including effecting the reproduction of fish, reduction in biodiversity, and, dysmorphia in the maturation of organisms (Jobling *et al.*, 1998; Sonnenschein and Soto, 1998; Beketov *et al.*, 2013; Malaj *et al.*, 2014; Capaldo *et al.*, 2018). Furthermore, compounds with similar biological activity have been shown to have synergistic effects, increasing the effect when compared to an individual pollutant (Relyea, 2009; Wang *et al.*, 2017).

The impact of research on organic micropollutants has led to legislation in the EU, as part of the water framework directive, to monitor 33 priority pollutants in surface water across Europe (Directive 2000/60/EC). Since this legislation was introduced there have been numerous research studies that found other biologically active compounds which are not part of the water framework directive but in environmentally relevant concentrations can adversely impact upon organisms in the aquatic ecosystem (e.g. Oetken *et al.*, 2005; Fong and Hoy, 2012; Gavrilescu

et al., 2015; Jones *et al.*, 2015; Barkham, 2018; Capaldo *et al.*, 2018). The consequences of these emerging pollutants led to multiple high-profile media articles (e.g. Sella and Lorna, 2014; Barkham, 2018). This demonstrates a growing public concern about the effect that biologically active synthetic compounds are having on our environment.

1.2 Origin of micropollutants in the aquatic ecosystem

Sources of micropollutants in the aquatic environment can be associated with two transportation mechanisms:

(i) Point sources, which have a clearly defined origin, such as an outflow pipe.

(ii) Diffuse sources, where organic pollutants found in the surrounding environment are transported via hydrological processes into the aquatic ecosystem.

Once anthropogenic compounds enter into the environment and mix with the existing DOM in the aquatic system it is difficult to determine which compounds are synthetic pollutants and which are derived from natural sources of organic matter (Ruff *et al.*, 2015; Hollender *et al.*, 2017). This is further complicated by the fact that DOM is highly dynamic and changing at a molecular level undergoing physical, chemical and biological processing (Obernosterer and Benner, 2004; Gallampois *et al.*, 2013; Haddad and Kümmerer, 2014; Mangal *et al.*, 2016). Therefore, a known compound entering the environment may be converted to a number of transformation products by instream processing (Allard *et al.*, 1994; Haddad and Kümmerer, 2014). These transformation products may still be biologically active and therefore it is important to identify them so their bioactivity and toxicity can be studied. Therefore, DOM in the aquatic environment comes from a variety of origins and its composition depends on not only the sources, but also how it is treated and utilised in the urban water cycle.

1.2.1 Urban water cycle

In the UK approximately a third of tap water is abstracted from aquifers whilst the rest comes from surface water, including lakes, reservoirs and rivers. Figure 1.2 is a depiction of the urban water cycle. Water is abstracted from the fresh water source and is treated for distribution to homes.

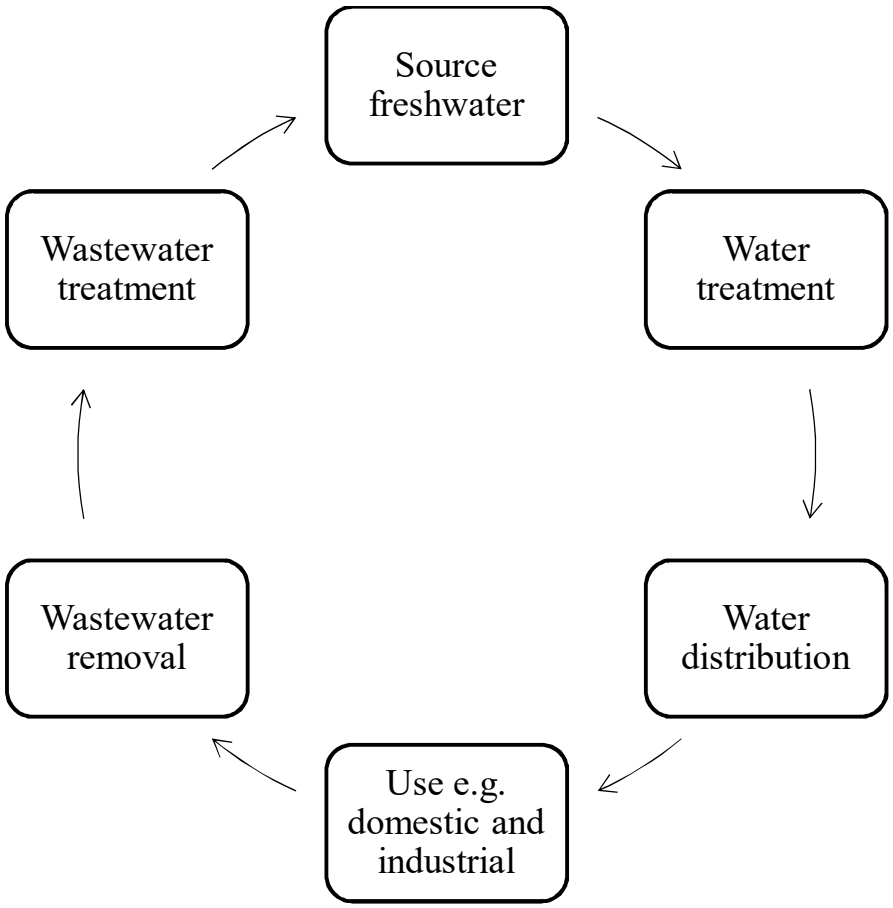


Figure 1.2. A depiction of the urban water cycle.

In the England and Wales 15,641 MLd^{-1} is distributed to 57.7 million people (Drinking Water Inspectorate, 2001). Treatment methods of natural fresh water can be split into 3 stages:

(i) Removal of particulates and algae using settling, screens, sand and gravel beds.

(ii) Chemical treatment which varies depending on the source of fresh water but can include: pH control, removal of metals such as iron and manganese, and taste or odour removal using activated carbon or aeration.

(iii) Disinfection. The most common method being chlorination, however, UV disinfection is used in smaller treatment plants (Drinking Water Inspectorate, 2018). The treatment of drinking water is not designed to remove natural DOM found in the environment before treatment unless it affects taste or odour. This has caused problems in drinking water production because the chlorination of the DOM in water has been shown to form chlorinated by-products which are potentially carcinogenic (e.g. Krasner *et al.*, 2006; Kim and Yu, 2007; Beggs *et al.*, 2009; Goslan *et al.*, 2009; Allard *et al.*, 2012).

This treated drinking water is supplied for use by a wide variety of different domestic, commercial and industrial processes. As a by-product of water use, wastewater will contain a range of inorganic and organic compounds, some of which could impact upon the aquatic environment. The majority of liquid waste is discharged into the sewer system. However, domestic and industrial works not connected to the sewage network can apply for an environmental permit to discharge wastewater into local surface water dependent upon an environmental assessment. This wastewater needs to be treated before being returned to the environment.

In the UK alone approximately 11 billion litres of sewage are produced every day and treated by approximately 9,000 sewage treatment plants then discharged back into the environment (DEFRA, 2002). The treatment processes used to treat sewage depends on the "sensitivity" of an aquatic system as set out in the EU urban wastewater directive (Directive 91/271/EEC; DEFRA, 2002). A "less sensitive area" is defined as an estuary or costal water that is not at risk of eutrophication. A "sensitive area" is defined if it meets one of three criteria:

(i) The area is at risk of eutrophication.

(ii) The drinking water sources could contain more than 50 mg L^{-1} of nitrate.

(iii) The water is in a protected area because of other designations under the EU water framework directive, such as bathing or shellfish waters (Directive 2000/60/EC).

The processing of wastewater at a sewage treatment plant before it is returned to the environment can be split into four parts (see Figure 1.3):

(i) Preliminary treatment which is the removal of large particulate matter.

(ii) Primary treatment which allows particulate matter to flocculate and settle out of the water column.

(iii) Secondary treatment which uses microbial degradation to breakdown biodegradable organic matter.

(iv) Tertiary treatment varies depending on the area of discharge. It is designed to mitigate any effect not moderated by the previous three methods on the aquatic ecosystem, such as phosphate removal, denitrification or additional filtration processes.

The UK has withdrawn from the EU designation of "less sensitive" areas meaning that the minimum requirement for sewage is higher in the UK than EU. The Government requires that

all sewage treatment plants in the UK must use a preliminary, primary and secondary treatment method (DEFRA, 2002, 2012).

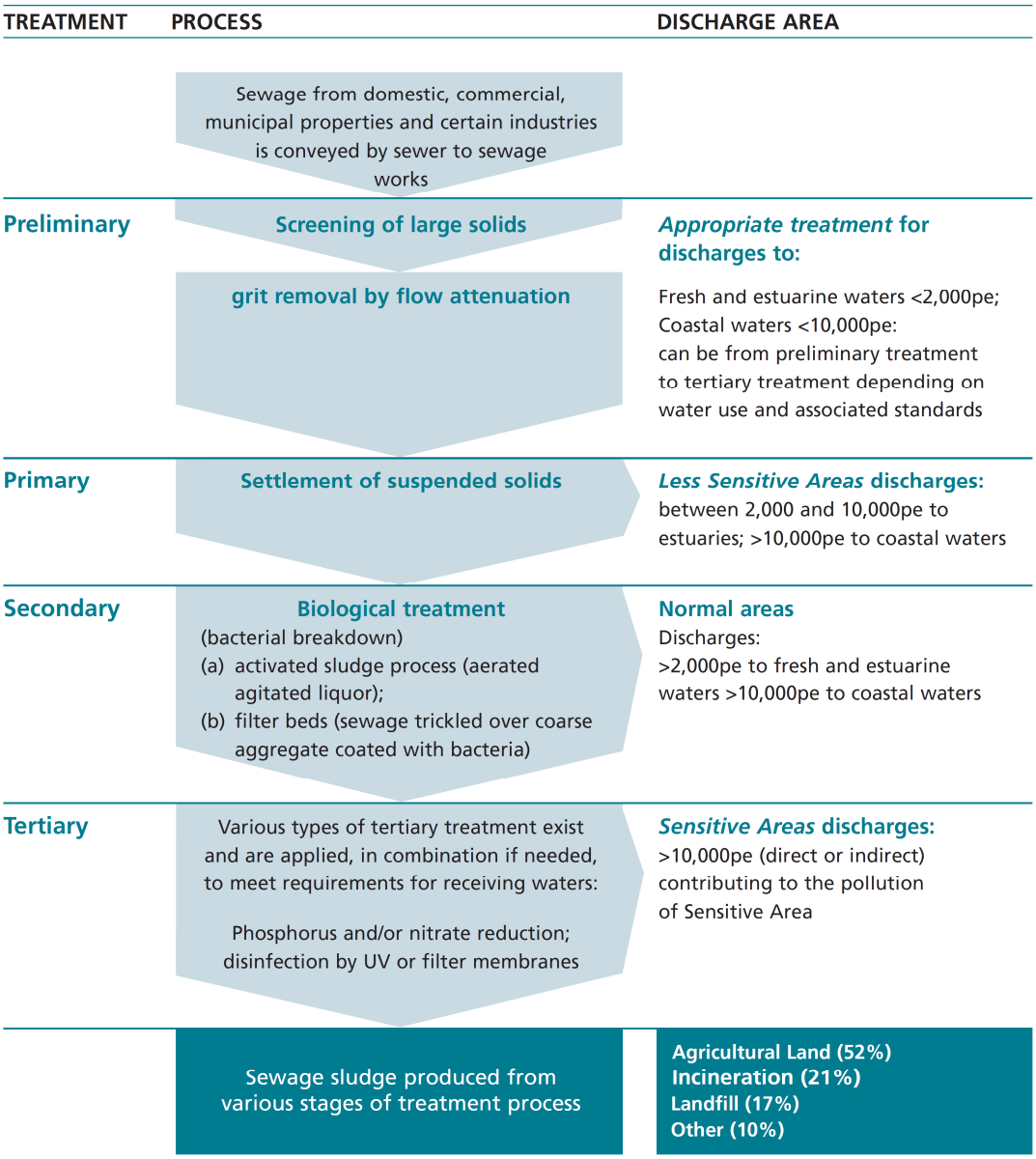


Figure 1.3. Diagram of the four stages of the sewage treatment process including discharge areas for sewage effluent adapted from (DEFRA, 2002). Population equivalent (pe) is used to describe the size of the wastewater discharge: pe is the amount of biodegradable matter in wastewater where the biochemical oxygen demand requires 60 g oxygen per day. pe does not reflect the actual size population but is based on the average amount of wastewater and organic matter produced by a person.

The biological treatment of sewage is the only part of the process which focuses on the removal of dissolved organic compounds and is designed to prevent eutrophication caused by high nutrient loads entering the local environment. However, it has been shown that biologically stable micropollutants persist through these processes and are discharged into the environment. These include, pharmaceuticals, illicit substances and surfactants (e.g. Longwell and Maniece,

1955; Loraine and Pettigrove, 2006; Mills *et al.*, 2007; Gros *et al.*, 2009; Karolak *et al.*, 2010; Gonsior *et al.*, 2011; Goel and Kaur, 2012; Santos *et al.*, 2013; Baker *et al.*, 2014; Maida *et al.*, 2017). The sewage processed will be a mixture of domestic, industrial and commercial liquid waste produced by a specified area. Depending on the population and industries using the sewage treatment works the composition of the sewage will change and therefore so will the composition of the effluent after treatment. Similarly, the treatment processes used to degrade DOM have been shown to generate a number of different transformation products from anthropogenic compounds (Haddad and Kümmerer, 2014; Hollender *et al.*, 2017). The different treatment methods used will not only create different transformation products but also vary in the effectiveness of removing certain chemicals. Therefore, the composition of effluent discharged into the environment may be specific to a particular treatment works.

Some compounds have a detrimental impact upon the ecosystems at every trophic level. Endocrine disrupting chemicals for instance, caused sexual disruption in fish, cocaine damages the skeletal muscle of eels and some pharmaceutical agents bioaccumulate in food webs (e.g. Jobling *et al.*, 1998; Henry *et al.*, 2004; Oetken *et al.*, 2005; Li, 2013; Zenker *et al.*, 2014; Ford and Fong, 2016; Capaldo *et al.*, 2018). Once effluent is discharged into the surface water it is transported through the environment, contaminating other surface water and percolating into groundwater. However, in order to determine the impact of sewage outfall or other point source upon the natural ecosystem, it needs to be comprehensively characterised at the molecular level before the compounds can be quantified. Then assessments of ecotoxicity and regulation or mitigation of the effects can be introduced to protect the environment.

1.3 Current methods for analysis of dissolved organic matter

DOM is operationally defined as the organic content of water after being filtered through a 0.45 μm filter with anything retained being defined as particulate organic matter (Spitzy and Leenheer, 1991; Norrman, 1993; Brailsford *et al.*, 2017). However, this definition of DOM is not consistently applied in the literature. The pore size of filters ranges from 0.7 μm to 0.2 μm, however, the filtrate is collectively referred to as DOM (Longnecker and Kujawinski, 2011; Li and Minor, 2015; Lv *et al.*, 2016; Phungsai *et al.*, 2016). Furthermore, studies have shown that microbes and viruses are able to pass through 0.45 μm filters. Therefore these samples are not biologically inert (Norrman, 1993; Brailsford *et al.*, 2017). Indicates that the filtrate does not solely contain dissolved compounds from water, but additional colloidal and biological matter.

Due to the complexity and dynamicity of DOM a range of analytical methods are needed for monitoring its concentration and composition.

The collective properties of DOM are measured using both *in situ* and laboratory techniques. These include elemental analysis, Ultraviolet–visible (UV-vis) spectroscopy and fluorescence spectroscopy. These methods measure the combined properties of the composition of DOM and do not require preconcentration of DOM nor the removal of inorganic compounds prior to analysis. These analyses can be applied *in situ* and enable a high throughput of water samples in the laboratory. This allows the analysis of quantitative changes of DOM in the environment, the identification of temporal trends, and the concentration of point and diffuse pollution. Some insights into the qualitative changes of DOM can be determined but it is not possible to characterise the individual compounds by using these analytical techniques

Characterisation of DOM using nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) requires the extraction and preconcentration of DOM prior to analysis (Minor *et al.*, 2014). It is necessary to remove any inorganic salts as these can affect the analytical techniques used for these analyses. The preconcentration of DOM is essential to concentrate individual compounds which can be present in concentrations as low as < 1 ng L^{-1}. The extraction of DOM is carried out using a variety of different techniques including, liquid-liquid extraction, reverse osmosis and electrodialysis, ultrafiltration and solid phase extraction (SPE) (e.g. Serkiz and Perdue, 1990; Boyd-Boland *et al*., 1996; Sáenz Barrio *et al.*, 1996; Müller *et al.*, 2004; Liu *et al.*, 2010; Cortés-Francisco and Caixach, 2013; Minor *et al.*, 2014; Li *et al.*, 2017).

1.3.1 Elemental analysis of DOM

Accurate measurements of the concentration of dissolved, particulate, organic and inorganic forms of nitrogen, phosphorus and carbon are essential for understanding the biogeochemistry in any environment. These elements form the essential nutrients and understanding their concentration is necessary to determine nutrient loads in the environment. These methods have been useful to show the way that anthropogenic sources affect the environment.

Agriculture has been found to be the key anthropogenic driver increasing nutrient loads in rivers. This is due to irrigation and precipitation mobilising inorganic fertilisers and labile soil organic matter into streams and rivers (Heathwaite *et al*., 1996; Novotny, 1999; Pinckney *et al.*, 2001). However, sewage treatment plants have been found to be anthropogenic point

sources of these nutrients by creating a localised increased nutrient load downstream of the point of discharge into the environment (DEFRA, 2012; Haddad and Kümmerer, 2014). The concentration of these nutrients in the sewage effluent depends on the hydrological state of the water received by the plant and the effectiveness of the treatment process being used to remove these nutrients. The sewage outfalls have the potential to have catastrophic effects on the immediate surrounding aquatic environment. Increased nutrient loads in the environment are a key cause of eutrophication.

These findings have affected how sewage is treated and, in the EU, have led to the implementation of the wastewater directive Figure 1.3 (Directive 91/271/EEC; EEC and European Union, 1991). Secondary treatment is designed to remove the amount of biologically degradable organic matter to reduce nutrient loads into the environment (DEFRA, 2002).

1.3.2 Ultraviolet–visible spectroscopy (UV-vis)

UV-vis is used in the analysis of both natural water and wastewater. The absorption of UV light at specific wavelengths has been shown to correlate with specific properties of DOM. A study by Dobbs *et al*. (1972) found a correlation between the total organic carbon content of water and the absorption of the wavelength 254 nm in both natural water and wastewater. This research lead to other absorption wavelengths which have been found to also correlate to the concentration of DOM including 350 nm and 440 nm (Spencer *et al*., 2012). These correlations have been used to develop *in situ* sensors which measure the specific UV absorbance (SUVA) typically at the wavelength 254 nm. These sensors can be used to provide real time *in situ* approximations of the changes in concentration of DOM in an aquatic system. UV-vis analysis of water samples also used the spectral slopes between different wavelengths and have been shown to correlate with specific properties of DOM. For example, the spectral slope 250 nm : 365 nm has been found to correlate with the concentration of aromatic material in a water sample and be inversely proportional to molecular weight (De Haan *et al.*, 1982; Dalzell *et al*., 2009).The caveats of using UV-vis to analyse DOM is that only compounds which contain functional groups which absorb UV light contribute to the signal recorded. As, these compounds only make up a fraction of DOM, this technique is not representative of all DOM.

1.3.3 Fluorescence spectroscopy

Florescence spectroscopy is used extensively both in the monitoring of DOM in wastewater and the natural environment. Excitation and emission spectra from florescence spectroscopy

infer source and structural changes in DOM using changes in the peak intensities. These multidimensional spectra can be used to identify peaks using a peak picking or parallel factor analysis, which change when comparing different environments or point sources (Ishii and Boyer, 2012; Carstea *et al.*, 2014). Peaks in the fluorescence spectra have been correlated to natural fluorophores in the environment including, lignin, amino acids and polyaromatic hydrocarbons (Fellman *et al*., 2010; Ishii and Boyer, 2012; Wünsch *et al*., 2015; Carstea *et al.*, 2016). Despite, fluorophores providing an indication of the structure of a compound, multiple compounds share the same excitation and emission wavelengths (Wünsch *et al*., 2015; Carstea *et al.*, 2016; Stedmon and Cory, 2018). Therefore, the specificity of these florescence peaks cannot be used to unequivocally identify individual compounds.

Despite the lack of specificity of peaks, spectrofluorometry has the potential to be used in the on-line monitoring in the processing of sewage. The depletion of peaks in the florescence spectra have been correlated to the biochemical oxygen demand of sewage (Reynolds and Ahmad, 1997; Ahmad and Reynolds, 1999). This improves upon current methods for determining the biochemical oxygen demand. Current methods require the incubation of sewage samples. Toxic components in sewage can kill these microbial cultures (Bourgeois *et al.*, 2001). Nevertheless, florescence has the advantage of providing *in situ* measurements of microbial degradation in the secondary treatment process of sewage (Reynolds and Ahmad, 1997; Ahmad and Reynolds, 1999). Florescence spectroscopy has also been shown to be useful in tracing the signals of sewage effluent into the natural environment (Pfeiffer *et al.*, 2008).

1.3.4 Extraction of DOM for analysis using NMR and MS

MS and NMR require DOM to be isolated and pre-concentrated prior to analysis. As the concentration of individual compounds can be < 1 ng L^{-1} (Mook and Tan, 1991).There are four commonly used extraction methods to isolate and preconcentrate DOM; ultrafiltration, reverse osmosis coupled with electrodialysis (RO/ED), liquid-liquid extraction (LLE) and solid phase extraction (SPE) (Minor *et al.*, 2014).

Ultrafiltration is a method of extraction which uses a molecular weight cut off membrane to extract a higher molecular weight DOM fraction (Müller *et al.*, 2004). The water and salts are allowed to pass through the membrane and organics, typically above 1000 Da, are retained on the membrane to be extracted for further analysis (Simjouw *et al*., 2005). Typically, ultrafiltration extracts a large proportion of DOM and does not use harsh chemical manipulations, such as acidification. However, this technique is limited to the analysis of high

molecular weight DOM. Furthermore, additional desalination techniques are required to fully remove salts prior to analysis.

RO/ED is used predominantly for large scale isolation and preconcentration of both marine and freshwater DOM. It reportedly has the highest recoveries (>90%) of DOM (Serkiz and Perdue, 1990; Fritzmann *et al.*, 2007; Green *et al.*, 2014). This technique creates a pressure difference across a permeable membrane retaining organic and inorganic compounds and removing water. The inorganic components are removed by electro dialysis. A further concentration step may be required such as lyophilisation to concentrate the samples further. The cost and set up of RO/ED extraction means it is not a practical solution to analysing the composition of DOM from different sites and point sources (Green *et al.*, 2014).

LLE of water samples has also been used for targeted analysis and monitoring of particular compounds including pesticides, sulphonamides and polyaromatic hydrocarbons (Sáenz Barrio *et al.*, 1996; Liu *et al.*, 2010). LLE is the standard technique for monitoring some priority pollutants in the EU water framework directive and EU drinking water directive (Directive 2000/60/EC, Directive 98/83/EC). However, LLE is labour intensive for the processing and extraction of multiple samples therefore the water quality industry has replaced many LLE protocols with SPE.

SPE is the most commonly used extraction method for DOM, it provides a single extraction and preconcentration step. SPE uses a solid phase, typically a polymeric or silicate phase that preferentially retains organic compounds on the solid phase, letting salts and water pass through the cartridge (Liška, 2000; Sabik *et al.*, 2000; Buszewski and Szultka, 2012; Raeke *et al.*, 2016). The organic extract can be eluted from the cartridge by switching to an organic solvent or changing the pH. The solid phase can be used to extract a broad spectrum of DOM or a particular phase can be chosen to optimise the extraction of specific compounds of interest (Zhou *et al.*, 2012; Cao *et al.*, 2015; Raeke *et al.*, 2016). SPE is a preferred technique because of its reproducibility, extraction efficiency, low cost and the ability to automate the extraction process enabling the high throughput of samples (Peček *et al.*, 2013; Green *et al.*, 2014; Lababidi and Schrader, 2014). However, it has been shown that different cartridges extract different fractions of DOM depending on the specificity of the solid phase used (Li and Minor, 2015; Raeke *et al.*, 2016; Arellano *et al.*, 2018). Furthermore, the acidification or basification of the water sample prior to extraction could alter the composition of compounds within DOM (Minor *et al.*, 2014).

All of these extraction methods only extract a fraction of DOM in a water sample. Therefore, depending not only on the extraction method used but also the membrane, solvent or solid phase will determine the composition of the fraction of DOM analysed (Flerus *et al.*, 2011; Green *et al.*, 2014; Minor *et al.*, 2014; Li and Minor, 2015; Li *et al.*, 2017).

1.3.5 Nuclear magnetic resonance spectroscopy (NMR)

NMR has been used to analyse extracts of DOM. Comparing different NMR spectra, at different sites or temporal environmental events can provide insight into the change in composition of DOM (Cortés-Francisco *et al.*, 2014; Zhang *et al.*, 2014; Hertkorn *et al.*, 2016; Li *et al.*, 2016). However, due to the complexity and heterogeneity of DOM there is a large amount of overlap between signals in the NMR spectra. 1D NMR experiments cannot resolve individual peaks therefore a range of 2D and 3D experiments are often applied to the analysis of DOM (Kaiser *et al.*, 2003; Kim *et al.*, 2003; Hertkorn *et al.*, 2013). Using NMR, the predicted shifts characteristic of a compound class or functional group, can be used to determine if this functionality is present in the extract DOM analysed. NMR provides insight into the functionality of compounds and molecular composition of DOM (Kaiser *et al.*, 2003; Hertkorn *et al.*, 2013, 2016). However, previous studies using NMR have proposed but not been able to confirm the identity of individual molecules from DOM (Kaiser *et al.*, 2003; Kim *et al.*, 2003; Zhang *et al.*, 2014; Hertkorn *et al.*, 2016). Therefore, studies of DOM using NMR are often used in conjunction with high-resolution MS (HRMS) to provide further insight into the composition of DOM at a molecular level.

1.3.6 Mass spectrometry (MS)

The characterisation of individual compounds found in DOM are identified and quantified using MS. The analysis of DOM using MS can be split into two experimental designs which are used to investigate the composition of DOM. These are targeted and untargeted analysis (Schymanski *et al.*, 2015).

1.3.6.1 Targeted analysis using MS

Targeted analysis has been used to identify compounds in both wastewater and the natural aquatic environment. Targeted analysis focuses on identifying and quantifying a list of known or suspected compounds. (Rodrigues *et al.*, 2007; Gros *et al.*, 2009; Santos *et al.*, 2013). This methodology uses optimised extraction methods to isolate and pre-concentrate the target

analytes. Using chromatographic separation including gas chromatography (GC) and high-performance liquid chromatography (HPLC) prior to MS analysis, these compounds can be quantified using known reference standards.

Targeted analysis studies predominantly focus on a particular class of compounds such as, pharmaceuticals, personal care products or food additives and a determination is made of their concentrations in the water (Ying, 2006; Kosonen and Kronberg, 2009; Sheng *et al.*, 2014). Comparing the concentration of compounds and their loading in the environment, allows the effectiveness of sewage treatment and the sources of the target compounds to be determined (Verlicchi *et al.*, 2012; Santos *et al.*, 2013).

An EU survey of 90 European wastewater treatment plants across 6 research centres conducted a targeted study of 161 organic chemicals. These chemicals belonged to a range of different compound classes including pharmaceuticals, veterinary drugs, personal care products, perfluoroalkyl substances, organophosphate flame retardants, pesticides and industrial chemicals (Loos *et al.*, 2012). The list was compiled from previous research studies of wastewater treatment plants. From this study it was found that 37 compounds were present in >85% of the 90 sewage effluent samples analysed. Thirty-one compounds were found in less than 10% and 28 compounds were not detected in any of the 90 sewage treatment works. Furthermore, the compounds concentrations showed a high level of variance between the different sewage effluent samples. For example, Acesulfame K which is a low-calorie artificial sweetener was detected in 93% of sewage effluents but the concentration ranged from 2.48 mg L^{-1} to 110 ng L^{-1} and a median concentration of 14.3 μg L^{-1}. This shows that many compounds consistently persist through current wastewater treatment processes. However, the composition of sewage effluent varies as do the concentration of individual compounds.

Targeted studies of the quantity of compounds in wastewater show seasonal, weekly and diurnal variation. A study by Baker *et al*. (2014) showed that the concentration of recreational illicit drugs in untreated sewage increases during the weekend when compared with the average concentration during the week. Furthermore, pharmaceuticals such as antihistamines have been shown to change seasonally in concentration correlated to higher pollen counts in the summer and lower counts in the winter (Kosonen and Kronberg, 2009; Golovko *et al.*, 2014). This shows that there is a temporal variability in the composition of wastewater and as a result the variation of the concentration of these compounds in the environment.

The limitation of targeted analysis is that it requires a predetermined list of known compounds. Targeted analysis will only investigate these selected compounds and exclude other compounds originating from a point source or the environment. A study by Haddad & Kummerer (2014), for instance analysed the photo-degradation of ciprofloxacin which is a broad-spectrum antibiotic commonly used in human and veterinary medicine. It showed 11 degradation products identified still retained the core quinolone structure thought to be responsible for the biological activity in the molecule. This study shows the complexity of a single molecule and a degradation pathway, further exemplifying the complexity of DOM in aquatic systems. Many studies analysing the compound ciprofloxacin in surface water and sewage outfalls have not been able to detect it, however, they have not targeted the degradation products (Gros *et al.*, 2006; Santos *et al.*, 2013; Ruff *et al.*, 2015).

Even though the treatment of sewage and drinking water are not specifically designed to remove organic pollutants these treatment processes can still degrade or form transformation products. Therefore, a reduction in concentration caused by a treatment process may not be indicative of the removal of a compound of interest but instead the target analyte may have undergone biological, physical or chemical transformation.

In summary there are two shortcomings with targeted analysis. Firstly, this approach is not a comprehensive representation of all compounds present in DOM. Secondly, a reduction in concentration of a compound is not necessarily indicative of its removal but may be a result of transformation.

1.3.6.2 Untargeted analysis using MS

Modern developments in soft ionisation, and HRMS have revolutionised the analysis of complex mixtures (Dass, 2006; Hawkes *et al.*, 2016). Liquid based soft ionisation techniques such as electrospray ionisation (ESI) and atmospheric pressure photoionisation (APPI) ionise a compound by the addition of an adduct or the removal a proton to form an intact molecular ion. The devolvement of Fourier transform ion cyclotron resonance MS (FTICR-MS) and Orbitrap MS which have high mass-resolving power and mass accuracy, can with high accuracy and precision, measure the mass of an ion (Hawkes *et al.*, 2016). These advances in soft ionisation and HRMS allow complex mixtures to be directly injected into an MS, and thousands of individual molecular ions to be determined simultaneously with or without prior chromatographic separation.

First of all the development of HRMS has led to two different approaches to the untargeted analysis of DOM. Direct infusion MS (DI-HRMS), which has led to studies comparing the qualitative changes in DOM between different environmental settings and extraction methods (e.g. D'Archivio *et al.*, 2007; Green *et al.*, 2014; Melendez-Perez, Martínez-Mejía, *et al.*, 2016; Li *et al.*, 2017). Secondly, HPLC-HRMS which aims to identify and quantify compounds which are ecologically significant (e.g. Gonsior *et al.*, 2011; Hug *et al.*, 2014; Schymanski *et al.*, 2014; Ruff *et al.*, 2015)

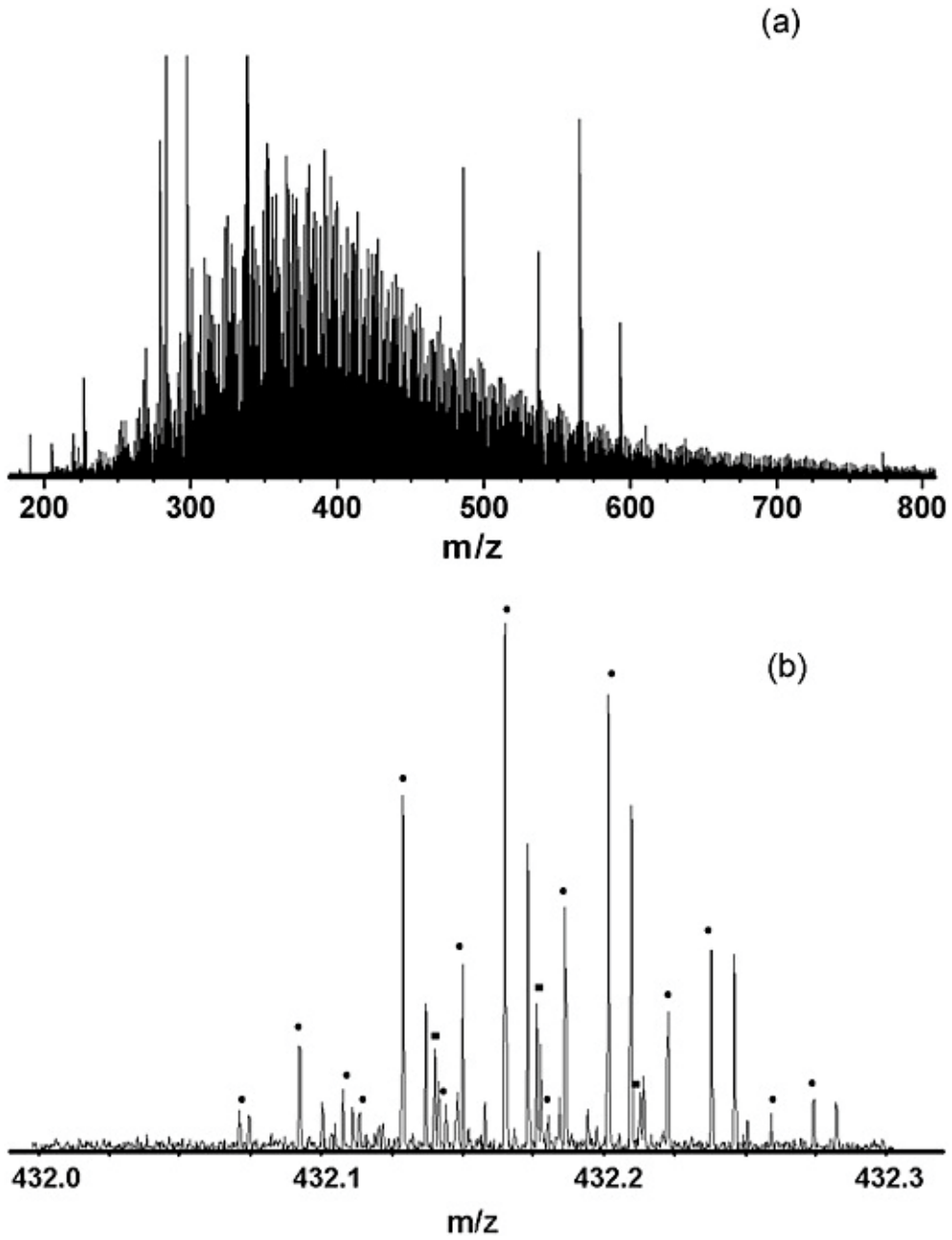


Figure 1.4. (a) Direct infusion positive ion APPI-FTICR spectrum of Varien Bond Elut PPL SPE extracted DOM from Caloosahatchee River water mass range 150 to 800 *m/z* (b) Direct infusion positive ion APPI-FTICR spectrum of Varien Bond Elut PPL SPE extracted DOM from Caloosahatchee River water mass window 423.0 to 432.3 *m/z* (from Osborne *et al*., (2013))

The DI-HRMS spectrum of SPE extract of riverine DOM presented in Figure 4.1, shows a complex mixture of molecular ions. Within a *m/z* 0.3 mass range there are clearly multiple fully resolved ions. Each could be a unique compound within the DOM extract. The complexity of DI-HRMS spectra makes them difficult to compare. To determine the compositional differences between DI-HRMS spectra previous research studies, have predominantly characterised ions in the spectra by assigning molecular formulae. The molecular formulae are then represented on van Krevelen diagrams in order to compare the compositional differences (Figure 1.5) (e.g. Flerus *et al.*, 2011; Osborne *et al.*, 2013; Gonsior *et al.*, 2014; Dubinenkov *et al.*, 2015; Mosher *et al.*, 2015). However, there is an emerging number of research studies using multivariate statistics to determine if there is a difference between environmental settings or extraction methods (e.g. Chen *et al.*, 2014; Dubinenkov *et al.*, 2015; Li and Minor, 2015; Li *et al.*, 2017).

1.3.6.2.1 Molecular formula assignments for HRMS spectra

Molecular formulas are assigned to the observed ions in a spectrum, using rule-based molecular formula calculations. Current molecular formula assignments of ions in DOM extracts in DI-HRMS are often based on two research studies Koch *et al*. (2006) and Herzsprung *et al*. (2014). These studies both outline rules for assigning the molecular formula of DOM. However, the rules used to assign the molecular formula to ions in DI-HRMS spectra of DOM extracts can vary from study to study (e.g. Cortés-Francisco and Caixach, 2013; Osborne *et al.*, 2013; Chen *et al.*, 2014; Harris *et al.*, 2015). All studies include carbon, oxygen and hydrogen. However, the inclusion of heteroatoms for example of nitrogen, phosphorous, chlorine and sulfur vary from study to study (Cortés-Francisco and Caixach, 2013; Osborne *et al.*, 2013; Chen *et al.*, 2014; Harris *et al.*, 2015; Phungsai *et al.*, 2016). Furthermore, isotopes of ions are observed in DI-HRMS spectra, however, few studies use ^{13}C in formula assignments (Cao *et al.*, 2015; Mangal *et al.*, 2016). In addition, alkali metal adducts, such as $[M+Na]^+$ and $[M+K]^+$, are commonly formed when using soft ionisation, which are included in the formula assignment of some studies but not in others (Kujawinski and Behn, 2006). A study by Koch *et al.* (2007) showed that increasing the number of heteroatoms resulted in an exponential increase in the number of possible molecular formulae for a single ion. This shows the difficulty of assigning a molecular formula to a single ion observed in a DI-HRMS spectrum. This lack of consistency of rules in the literature for the assignments of molecular formulae makes it difficult to compare the composition of DOM across studies. Furthermore, because compounds are not identified in these studies it is difficult to validate the molecular formula assignments.

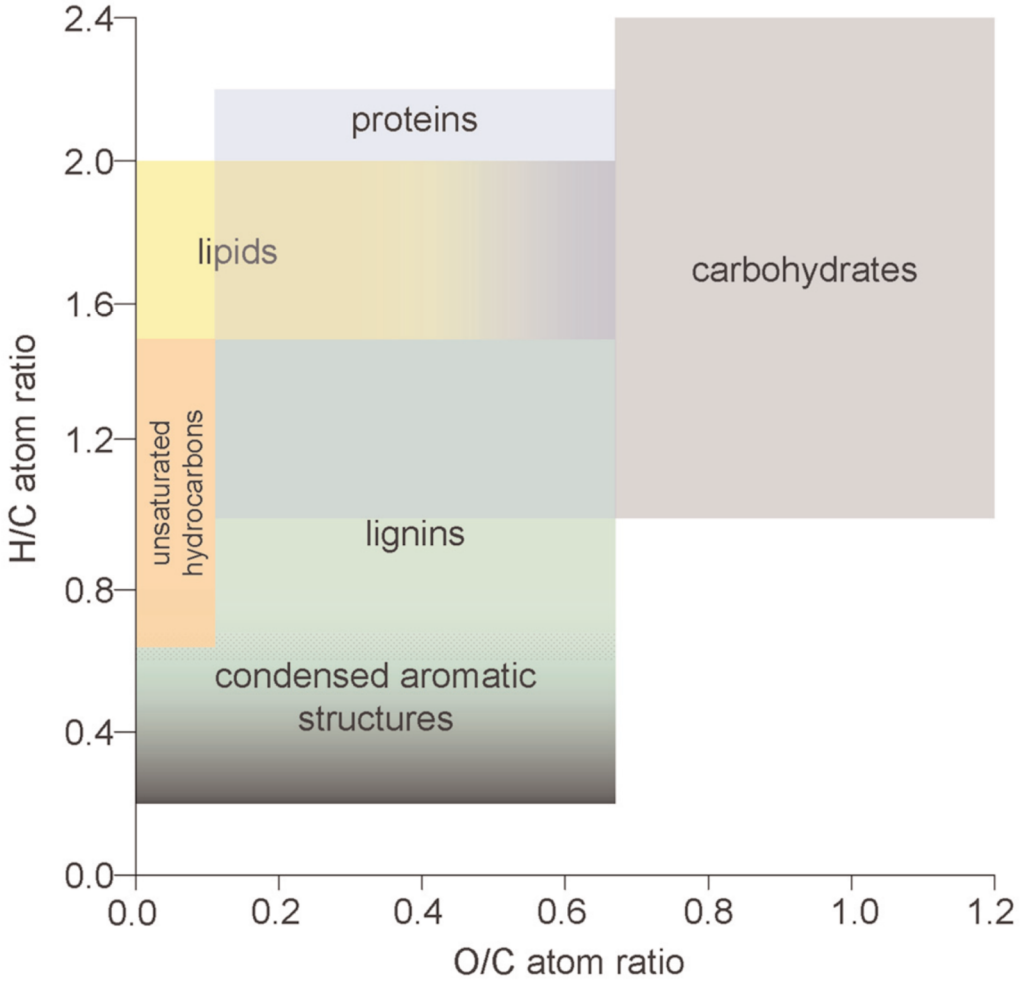


Figure 1.5. Model van Krevelen diagram showing the H/C and O/C areas used to classify natural products in DOM (adapted from Hockaday *et al.*, 2009).

However, by using molecular formula assignments, compositional changes between DI-HRMS spectra of different DOM SPE extracts can be visualised using van Krevelen diagrams (e.g. Sun *et al.*, 1997; D'Andrilli *et al.*, 2015; Mosher *et al.*, 2015; Herzsprung *et al.*, 2017). Using the ratios of carbon:hydrogen (H/C) and carbon:oxygen (O/C) of the molecular formulae assigned to ions allows the comparison of DOM SPE extracts. Furthermore, elemental ratios of a molecular formulae can be used to classify an ion to a compound class as shown by Figure 1.5 (Osborne *et al.*, 2013; Kellerman *et al.*, 2015; Li and Minor, 2015). However, the interpretation of a van Krevelen diagram relies on the correct assignment of molecular formulae. Incorrect assignment can lead to the inaccurate analysis of changes to the composition of DOM extracts.

Kendrick mass defect plots have been used to identify molecular series in DI-HRMS spectra of DOM extracts. Kendrick mass defect plots use common mass differences seen between

series of compounds such as -CH_2- to determine ions which may be part of an homologous series of compounds (e.g. Cortés-Francisco and Caixach, 2013; Gonsior *et al.*, 2014; Herzsprung *et al.*, 2014). However, Kendrick mass defect plots using thousands of ions in a DI-HRMS spectrum make it difficult to identify individual ions in a series.

1.3.6.2.2 Multivariate statistical analysis of HRMS spectra

Multivariate statistical methods, including principal component analysis (PCA) and hierarchical cluster analysis have recently been used to analyse the differences in the composition of DI-HRMS spectra, comparing SPE extraction methods and the composition of DOM at different environmental settings (Dubinenkov *et al.*, 2015; Melendez-Perez *et al*., 2016; Li *et al.*, 2017). For example, a study by Dubinenkov *et al*. (2015) analysed water extracts from across the Lena Delta, an Arctic river catchment. The study compared DI-HRMS spectra of SPE extracts from riverine, estuarine, lake/creek and ice complexes. It showed that by using PCA and cluster analysis that the organic molecular composition of the different extracts clustered depending on where in the aquatic system the water was collected from. However, these studies have not focused on anthropogenic pollution.

HRMS has been used to identify micropollutants from an untargeted perspective originating from sewage outfalls and in surface water. However, due to the complexity of sewage outfall and DOM in the natural environment it is not feasible to comprehensively characterise all compounds in an environmental setting or sewage outfall. Therefore, a variety of different methods are used to determine which ions in a mass spectrum to identify.

Gonsior *et al*. (2011) studied the difference in composition of an SPE extract from the effluent of a sewage treatment plant in Orange County California with that of a Suwannee River DOM reference RO/ED extract from Florida. The effluent from the two different treatment processes from the sewage treatment plant in California were blended. The DI-HRMS spectra in negative ion mode were measured for the blended effluent SPE extracts and the Suwannee river DOM and then compared. By assigning molecular formula it was found that there were more sulfur containing molecular ions in the effluent extract than in the reference extract. These formed an homologous series differing by CH_2. These compounds were identified as anionic surfactants, including linear alkyl benzene sulfonates, dialkyl tetralin sulfonates and sulfophenyl carboxylic acids. These compounds are known and are routinely identified in wastewater treatment plants from the targeted analysis studies of wastewater (Müller *et al.*, 2004; Ying, 2006). Gonsior *et al*. (2011) showed that the elemental composition of the DI-HRMS spectra could be used to

show differences between different DOM extracts. However, this study exclusively looked at the differences in ions in the DI-HRMS spectra which contained sulfur and excluded other potentially important differences between the reference and sewage effluent DOM. Furthermore, this study compared samples extracted using different methods. It is well known that different extraction methods extract different compounds from DOM (Sáenz Barrio *et al.*, 1996; Green *et al.*, 2014; Minor *et al.*, 2014; Li *et al.*, 2017).

A study by Hug *et al.* (2014) focused on a predetermined list of compounds suspected to be present in wastewater and the non-targeted identification of micropollutants. This study extracted the wastewater effluent over 6 days and used HPLC-HRMS to analyse the extracts. Peaks only found in all 6 days were retained for further analysis. To further reduce the number of compounds, only ions for which a molecular formula containing Cl, Br, S or N were retained for further analysis. Furthermore, only ions with an intensity value > 3 x 10^6 were retained for further analysis. The study stated that it assumed that all adducts were $[M+H]^+$. After molecular formula assignment the results were searched through Chemspider database with an average of 180 hits per molecular formula. This study excluded results with less than 10 data sources to reduce the number of hits. The study states that this is a drawback of the method as many emerging or novel compounds would be removed. This study fragmented 30 ions which had been highlighted during this process and used MetFrag fragmentation prediction software to compare the recorded product ion spectra to the predicted product ion spectra of compounds returned from Chemspider database. Through this procedure 4 additional compounds were tentatively identified. This research states that the size of a peak used in the initial filtering step does not necessarily mean that a compound is not environmentally relevant. This is because, an ions intensity in ESI will be an intrinsic property of how efficiently a compound is ionised and not directly related to its concentration.

Schymanski *et al.* (2014) performed a targeted, suspect and non-targeted study comparing the wastewater effluent from 10 treatment plants. Using the top 30 most intense peaks common to all the wastewater samples analysed by HPLC-MS analysis, the molecular formulae were assigned and searched through the Chemspider database. MetFusion was then used to predict the fragmentation of the results from Chemspider. One non-target compound was identified, 3-benzothiazole-2-sulfonate, which is a transformation product of 2-mercaptobenzothiazole. Similar to the study by Hug *et al*. (2014) this study has the same limitation in that the size of a peak is not necessarily an indication of concentration or

environmental importance. This study also stated a need for more extensive spectra libraries for environmental samples similar to those used in metabolomics.

Ruff *et al.* (2015) conducted a predominantly targeted study of micropollutants found in the extracts of surface water from the river Rhine. The Rhine is approximately 1233 km long and has a catchment area of 200,000 km^2 in six countries. Water sampling times were staggered based on the predicted flow of the Rhine from Lake Constance (Switzerland) to Bimmen-Lobith 865 km downstream to sample the same body of water. This study used reverse phase HPLC-HRMS to analyse SPE extracts for a target list of 302 compounds of which 128 target compounds were found. Ruff *et al.* (2015) stated that due to the large amount of data an appropriate prioritisation method was essential to select the focused identification of non-targeted compounds in surface water. To determine which ions to identify, Ruff *et al*. (2015) used an assigned molecular formula to highlight chlorinated compounds of which 2 additional compounds were identified. In addition, a systematic approach focusing on the top 150 components with the largest peak area were compared across each water extract and any components found in the top 150 components upstream were removed from the next downstream site. After a comparison of all the sites 126 emerging components were checked for peak shape and against the blank extracts. This left 57 components of emerging substances of which 3 were identified.

The use of this systematic data reduction using the presence and absence of a peak comparing the composition of the sites hundreds of km apart, does not account for the multiple sources of micropollutants between those sites. The staggered approach to water sampling assumes that the water is moving as a uniform "package" downstream. However, this does not consider the diffusion of pollutants in the river, variation in flow and changes in hydrological conditions. Therefore, how comparable these datasets are is difficult to determine. The use of a molecular formula-based approach to determine which compounds to analyse excludes all other compounds which do not have a heteroatom to differentiate them.

All previous untargeted analysis studies state the necessity to prioritise particular ions to identify from these large mass spectral datasets (Ruff *et al.*, 2015; Hollender *et al.*, 2017). These studies use a variety of different methods including trying to identify the most intense ions, using the presence/absence of peaks or using the molecular formula to determine which compounds to identify. However, all these methods have drawbacks which are summarised below.

The intensity of a peak using soft ionisation is not necessarily directly related to concentration but is also related the efficiency with which a molecule is ionised under the ionisation settings and method chosen. Furthermore, the justification for how many peaks or the intensity which makes a peak significant is subjective to the analyst.

Using the presence or absence of a compound in a sample extract could be used to determine if there is a qualitative change between two sampling locations. However, this depends on the comparability of the water extracts being analysed. Furthermore, without understanding the sources of pollution or hydrology between the locations it would be difficult to determine the origin of a compound. In addition, this excludes changes in concentration between two locations, which could also indicate an additional source.

The assignment of molecular formula to determine whether a compound's elemental composition makes it a pollutant has disadvantages:

(i) Many micropollutants which have been identified have no heteroatoms which would differentiate a compound from the other compounds in DOM,
(ii) This requires the correct assignment of molecular formula which can be difficult for complex mixtures, as discussed above,
(iii) The heteroatoms which make a molecular formula of a compound important is subjective.

Therefore, a better method is needed to determine which ions to prioritise for identification, when, using mass spectrometry for untargeted analysis.

1.4 Summary

This Chapter outlines the importance of DOM in aquatic systems in the natural environment and the effects that anthropogenic pollution can have. In particular, the emergent understanding of the impact that micropollutants can have on the aquatic ecosystem system. However, understanding the effect micropollutants are having starts with their identification and understanding their sources in the environment.

HRMS, has demonstrated the complexity of DOM in the natural environment and point sources using both untargeted and targeted analysis to identify micropollutants in the environment.

Although several methods have been used, the complexity of the data generated makes it difficult to determine which compounds are micropollutants and which are naturally occurring DOM. As outlined above all current methods have their drawbacks. There, is a need to develop an alternative method to determine which compounds to focus on. As discussed, the use of multivariate statistical analysis of DI-HRMS spectra to differentiate between environmental settings and extraction methods, shows the potential for it to be used to analyse point sources in the environment.

All point sources have a defined point of origin and rivers flow in a single direction depend on the topography of the landscape. This creates a one-way transportation mechanism for DOM. Therefore, the DOM of a point source discharging into a river will be transported downstream. A comparison of the DOM in the water upstream, a point source and a downstream extract would create a comparative dataset. Statistics could then be applied to address the following questions:-

(i) Is the point source composition different to the river?
(ii) Does the point source affect the composition of the river downstream?
(iii) Which compounds are most significant for differentiating between the point source and riverine DOM?

1.5 Thesis aims and hypotheses

This thesis, therefore, aimed to develop a novel statistically driven untargeted analysis method to determine the difference in composition between point sources and the environmental setting they are discharging into. Furthermore, it aimed to determine the compounds which are discriminating and are being added to the environment by the point source.

The primary aims of this thesis were to:

(i) Determine using DI-HRMS analysis whether the composition of a point source is significantly different to the environmental setting it is discharging into. Then examine if these differences between DI-HRMS spectra can be used to highlight which ions are driving these changes as a data reduction step. (Chapter 3).
(ii) Develop an analytical method to identify compounds determined to be statistically significant. Then determine whether these compounds identified as statistically significant can be related to the published literature of compounds previously identified (Chapter 4)

(iii) Apply this method to different to study different point sources in different environmental settings of increasing complexity, in order to determine the applicability of this method to other point sources (Chapter 5 & 6)

(iv) Compare the compositional differences between environmental settings and point sources. (Chapter7)

The overall hypotheses are:

(i) Processed sewage effluent will be a highly complex mixture of compounds many of which will be overlooked in a targeted analysis study, but may still be of environmental relevance.

(ii) The molecular "fingerprint" of the extracted sewage effluent will be similar as similar products are likely to be used in UK households which will be discharged as sewage. Therefore these compounds are likely to be found in most sewage treatment works.

(iii) The molecular "fingerprint" of the upstream DOM SPE extract from different locations will show a similarly complex mixture of organic compounds. Previous literature shows a complex mixture of ions from riverine sources in different geographic locations (Osborne *et al.*, 2013; Minor *et al.*, 2014).

(iv) The composition of the sewage effluent DOM from the treatment works will differ significantly from the riverine DOM and will affect the composition of the river downstream of the treatment works.

(v) Through comparison of the sewage outfall's composition and the receiving aquatic environment, the DOM composition of not naturally occurring compounds can be determined.

(vi) That many of the compounds determined to have been contributed to the river from the sewage outfall will be the same as those compounds previously identified in published targeted studies i.e. known micropollutants.

(vii) As a result of the untargeted approach adopted herein several anthropogenic compounds will be detected that have been overlooked in previous studies i.e.. previously unknown micropollutants.

Chapter 2 will outline the methods and data processing used in subsequent chapters to test the aims and hypotheses stated above. Chapter 3 will introduce Chew Stoke sewage treatment

works going through in depth the DI-HRMS data of each of the SPE extracts providing novel data presentation to highlight the differences in their compositions. This Chapter will also outline the statistical methods used to compare the extracts and a novel unbiased data reductive step to determine the compounds to focus upon. Chapter 4 will explore analysing the extracts using HPLC-HRMS focusing on the identification of compounds coming from the sewage treatment works. Chapters 5 & 6 will demonstrate the applicability of this methodological approach to the analysis of point sources in catchment environments of increasing complexity. Chapter 7 summarises and compares the findings found in the previous chapters and outlines future work. Chapter 7 also discusses a preliminary study using passive samplers for an untargeted approach.

Chapter 2

Materials and Methods

Chapter 2. Materials and methods

2.1 Grab Sampling

All sample bottles were amber glass to minimise photodegradation and solvent washed with HPLC grade methanol (Rathburn Chemicals Ltd.) and left to evaporate before use. A glass bottle attached to a metal pole using metal wire was used to collect water from the river and sewage outfalls. The sampling equipment and storage bottles were rinsed five times with the water being sampled before the water sample (5.25 L) was collected. The polyethylene bottle cap was lined with a layer of furnaced aluminium foil to prevent contact between the water and lid of the bottle.

2.2 Solid phase extraction (SPE)

The water was divided into 1 L aliquots which were vacuum filtered using an all glass filter apparatus (47 mm, Merck Millipore) through glass fibre filters (0.5 µm, 47 mm, Advantec). Both the filter and filtration apparatus were furnaced before use (450 ºC, 4 h). An additional 20 ml of each water sample was filtered and retained for DOC analysis. The filtered water (1 L) was acidified to pH 2 using hydrochloric acid (Sigma–Aldrich TraceSelect, 30%) and extracted using Oasis Hydrophilic-Lipophillic Balance (HLB) solid phase extraction cartridges (400 mg bed mass, 60 µm particle size). The cartridges were conditioned using HPLC grade methanol (3 ml, Rathburn Chemicals Ltd.) and HPLC grade water (3 mL, Fisher Scientific) before the acidified filtered water (1 L) was extracted. After extraction, the cartridges were rinsed with HPLC grade water (3 mL, Fisher Scientific) and dried under vacuum for 30 min. The extracts were eluted from the SPE cartridges with HPLC grade methanol (6 x 1 mL, Rathburn Chemicals Ltd.) and dried under a steady stream of nitrogen. Dried extracts were dissolved in a known volume of HPLC grade methanol/water (1:1, *v/v*, Rathburn Chemicals Ltd., Fisher Scientific).

An aliquot of each extract (100 µL) was mixed to create a pooled quality control (QC) and an aliquot of each extract (50 µL) was removed and dried under a steady stream of nitrogen for DOC analysis. The pooled QC and all extracts were stored at –85 ºC until required for analysis.

2.3 Instrument analysis

2.3.1 Dissolved organic carbon analysis (DOC)

The dried 50 µL aliquots of the extracts were dissolved in water (20 mL, Milli q) before DOC analysis. Filtered water samples were analysed directly. All analyses were carried out using a Shimadzu TOC-L analyser using the non-purgeable organic matter (NPOC) method which is recommended by Shimadzu for analysis of environmental water samples. Freshwater samples contain a small amount of purgeable organic compounds, and therefore NPOC can be referred to as DOC (Lloyd *et al.*, 2012). The instrument method acidified the water sample to pH 2 and was sparged with air to remove any inorganic carbon. The sample was combusted at 680°C in the presence of a platinum catalyst. The resulting CO_2 is cooled and dehumidified before detection by a non-dispersive infrared detector (NDIR).

The area of the peak measured using NDIR of the CO_2 was be quantified by comparison to the peak areas of a range of standards with a known DOC concentration. TOC standard solution (1000 mg L^{-1}, Sigma-Aldrich) was diluted using water (Milli q) into standards solutions of 2, 4, 5, 6, 8, 10, 15 and 25 mg L^{-1}. All standards were run at the start of the analysis and a standard was run every five analyses to check for any changes in the accuracy of the instrument's performance.

The concentration of the organic carbon extracted was calculated using Equation 2.1; where DOC_{ex} is the total concentration of organic carbon in the SPE extract (mg L^{-1}), DOC_a is the determined concentration of organic carbon in the aliquot used for DOC analysis (mg L^{-1}), V_d is the volume of water used to dilute the dried aliquot (0.02 L), V_{ex} is the total volume of methanol/water used to dissolve the SPE extracts (µL) and V_a is the volume of aliquot taken from the total SPE extract (50 µL).

$$\boldsymbol{DOC_{ex} = DOC_a \times V_d \times \frac{V_{ex}}{V_a}} \qquad (2.1)$$

The extraction efficiencies were calculated using Equation 2.2; where EE is the extraction efficiency (%), DOC_{ex} is the total amount of organic carbon extracted and DOC_{rw} is the amount of organic carbon in the filtered water sample.

$$\boldsymbol{EE = \frac{DOC_{ex}}{DOC_{rw}} \times 100} \qquad (2.2)$$

2.3.2 Direct infusion mass spectrometry (DI-HRMS)

DI-HRMS was conducted in positive ion mode using an Orbitrap Elite Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Scientific) with a heated electrospray ionisation source (HESI). Extracts were directly infused at a rate of 5 μL min^{-1} into the HESI. The source settings are summarised in Table 2.1 and are based on the recommended settings outlined in the HESI-II probe guide (Thermo Scientific, 2009). The mass spectrometer was set to acquire in the mass range of *m/z* 150 to 2000 for 100 scans, and ions detected were recorded in profile mode using the mass resolution setting "240,000", the maximum injection time was set at 200 ms.

Table 2.1. HESI source settings for direct infusion mass spectrometry analysis

Source Voltage	3.0 kV
Source temperature	Off (approx. 60ºC)
Sheath gas	10 arb
Auxiliary gas	5 arb
Sweep gas	5 arb
Capillary temperature	275.06 ºC

SPE extracts were randomised before analysis. The pooled QC and calibration solution (Pierce™ LTQ ESI positive ion calibration solution, Caffeine, MRFA and Ultramark 1621) were analysed at the beginning and the end of the run as well as after every 5 SPE extracts. This was done to detect any analytical drift in the instrument's performance over the acquisition of the DI-HRMS spectra of all SPE extracts. Between each analysis the source was cleaned with a mixture of HPLC grade methanol/water (1:1, *v/v*, Rathburn Chemicals Ltd., Fisher Scientific) until the maximum injection time was reached (200 ms) to prevent any carry over between analyses. A solvent blank of methanol/water (1:1, *v/v*, Rathburn Chemicals Ltd., Fisher Scientific) was recorded between each analysis to ensure that there was no cross contamination.

2.3.2.1 Ion picking and alignment of DI-HRMS spectra

Files were converted from Thermo .raw to .mzML using MSConvert (Chambers *et al.*, 2012). All 100 scans were merged using an openMS spectramerger module in KNIME. This was done as the XCMS for the peak picking of DI-HRMS expects a single mass spectrum (Berthold *et al.*, 2009; Röst *et al.*, 2016). Ion picking and alignment was done using the XCMS package (v

1.52.0) in R (v 3.4.0) to create a data matrix of ion intensities aligned by mass (Smith *et al.*, 2006; Tautenhahn *et al*., 2008; Benton *et al*., 2010). Ions were picked using the mass spectral wavelet algorithm which is recommended for DI-HRMS spectra in XCMS.

Scan-to-scan variability of the accurate mass was assessed using the Orbitrap calibration standard (Pierce™ LTQ Velos ESI Positive Ion Calibration Solution) before and after the analysis of all the SPE extracts, using the mass check function in the Orbitrap software. The changes in the mass accuracy across the analytical run were assessed using the accurate mass of standard ions. These values were used to define the mass range used to align ions across different analyses using XCMS (v 1.52.0). The ion had to be present in 3 out of 5 of the DI-HRMS analyses.

Once the different DI-HRMS spectra of the SPE extracts were aligned, any missing intensity values for an ion were reduced by a second round of ion picking. This was carried out to determine if the ions found were in the other DI-HRMS spectra, but their intensity was below the signal to noise threshold when the ions were initially picked. This was done to minimise missing intensity values. All statistical analysis was then performed using Mass Profiler Professional (v 13.0, Agilent Technologies).

2.3.2.2 Pattern matching algorithm

The pattern matching algorithm was written in R (v 3.4.0). The exact masses of the ions found in Section 2.3.2.1 were subtracted to form a matrix of mass differences between all ions. The mass difference of interest was compared to the difference matrix and all differences within 10 ppm were highlighted. The ions which when subtracted resulted in the mass difference were tabulated. The series of ions were reconstructed by finding ions which were both the minuend and subtrahend.

2.3.3 High performance liquid chromatography high resolution mass spectrometry (HPLC-HRMS)

The SPE extracts were analysed using HPLC-HRMS using a Dionex Ulitmate HPLC system with a diode array detector (DAD) coupled to an Orbitrap Elite Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Scientific) with a HESI. Chromatographic separation used an Ace UltraCore Super C_{18} column (150 x 2.1 mm i.d., 25 Å particle size). The column was kept at a constant temperature at 50 ^{o}C. A gradient program with HPLC grade water (Fischer Scientific)

as mobile phase A and HPLC grade acetonitrile (Fischer Scientific) as mobile phase B both with a 0.1 % formic acid (Fisher Scientific) modifier was used. The flow rate was kept constant at 350 μL min^{-1}. The gradient program was as follows: 5% B for 1 min, 5% to 95% linear gradient for 30 min and 95% held for 5 min before returning to 5% in 1 minute and remaining at 5 % for 4 min. The DDA detector was set to record the wavelength 254 nm. This wavelength was chosen because absorbance at 254 nm has been correlated to the concentration of DOM in natural waters (e.g. Dobbs *et al*., 1972; Weishaar *et al.*, 2003).

The HESI settings are summarised in Table 2.2. All spectra were recorded at "120,000" resolution in positive ion mode for the mass range *m/z* 150 to 2000 in centroid. Between each analysis a solvent blank of HPLC water (Fischer Scientific) was run to ensure that there was no carry over between samples.

Table 2.2. HESI source settings for HPLC-HRMS analysis

Source Voltage	3.5 kV
Source temperature	80 ºC
Sheath gas	35 arb
Aux gas	10 arb
Sweep gas	10 arb
Capillary temperature	275 ºC

2.3.3.1 Peak picking and alignment of HPLC-HRMS data

Files from the HPLC-MS analysis were converted from Thermo .raw to .mzML using MSConvert (Chambers, 2012). Peak picking and alignment was performed using the XCMS (v 1.52.0) package in R (v 3.4.0) to create a data matrix of sample intensities aligned by mass and retention time (Smith, 2006; Tautenhahn, 2008; Benton, 2010). The method used for peak picking was the centWave algorithm which is recommended for peak picking and alignment HPLC-HRMS data. Peaks were picked above a signal-to-noise ratio of 10, the mass tolerance allowed was 10 ppm and a retention time tolerance range of 15 to 60 s. The peaks were then aligned across samples if the mass was within 0.002 Da and retention times overlapped by 10 s.

2.3.4 High performance liquid chromatography tandem mass spectrometry (HPLC-HRMS/MS)

Data dependant acquisition (DDA) method was used for the acquisition of MS^2 spectra for a target mass list of ions and their retention times. The HPLC method, DAD and source settings were consistent between the HPLC-HRMS and HPLC-MS/MS runs. The DDA used a pre-determined list of ions at specific retention times obtained from the previous HPLC-HRMS analysis. The method consists of 7 scan events. A full scan event recorded using the resolution mode "120,000" to identify the presence of a target mass ion. If a target mass ion was detected within the retention time range specified a series of 6 MS^2 scans were recorded in the Orbitrap using resolution mode "7000" at different CID energies of 10, 20, 30, 40, 50 and 60 eV.

2.3.4.1 Peak picking of HPLC-HRMS/MS data

HPLC-MS/MS files were converted from Thermo .raw to .mzML using MSConvert (Chambers, 2012). Peak picking was done using the XCMS (v 1.52.0) package in R (v 3.4.0). A data matrix of product ions and intensities was created corresponding to a specific precursor ion's mass, retention and the fragmentation energy (Smith, 2006; Tautenhahn, 2008; Benton, 2010).

Spectra were then compared with two databases mzCloud and MassBank. mzCloud was chosen because the database consists of fragmentation spectra predominantly recorded using Orbitrap mass spectrometers at multiple fragmentation energies. This not only allowed the matching of the product ions but also the intensities and fragmentation energies to be compared. Massbank was used as this is not only a large open source repository but is also the library used by the Norman Network which focuses on characterisation of molecules from environmental samples. mzCloud currently has no application programable interface (API) therefore, spectra were searched manually. Massbank has a freely available API and therefore a script in R (v 3.4.0) was written using the package Rcurl (v 1.95-4.10) to search product ion spectra via the Massbank database's searchSpectrumEx search function. All results for each production spectrum were compared to high scoring matches returned from the database to determine whether the product ion spectrum matched the database spectrum.

Chapter 3

Development of untargeted analysis of point source riverine DOM by direct infusion Orbitrap mass spectrometry (DI-HRMS): Using the River Chew and Chew Stoke sewage treatment works as a test site

Chapter 3. Development of untargeted analysis of point source riverine DOM by direct infusion Orbitrap mass spectrometry (DI-HRMS): Using the River Chew and Chew Stoke sewage treatment works as a test site

3.1 Introduction

Organic micropollutant components of DOM are of growing concern in aquatic ecosystems due to their potentially damaging effects (Petrie *et al.*, 2014; Gavrilescu *et al.*, 2015). Target compound analysis is the most commonly used approach to determining the nature and origins of organic micropollutants, however, this is limited to known compounds and their transformation products. In order to achieve an holistic assessment of the compounds potentially impacting on aquatic environments, a more comprehensive analytical approach is required to characterise the DOM organic micropollutant fraction.

The development of ESI-HRMS has revolutionised the analysis of complex mixtures, allowing extracts to be directly analysed by MS without prior chromatographic separation. The ionisation of intact molecules, usually as protonated adducts, $[M+H]^+$, and their mass analysis using instruments with high resolving power and mass accuracy means that each ion in a spectrum potentially corresponds to a unique compound (taking account of other adducts and isotopes). Application of this approach has revealed the extraordinary complexity and heterogeneity of DOM in the natural environment, evidenced in DI-HRMS spectra containing many thousands of resolved ions (Minor *et al.*, 2014; Li and Minor, 2015). The potential exists to investigate the identities of the various molecular species in such spectra, however, comprehensive identification of such an extensive array of unknown compounds would be a prohibitively time-consuming and laborious process. Therefore, the effective application of HRMS methodologies to the investigation of such complex environmental mixtures requires an alternative approach, involving the identification of the molecular species of most importance (Ruff *et al.*, 2015).

Pervious work in this area has recognised the challenge of distinguishing micropollutants from naturally derived DOM found in the aquatic environment. Both Gonsior *et al.* (2011) and Ruff *et al.* (2015) used a molecular formula-based approach to determine which compounds to identify, selecting ions with molecular formulae containing sulfur and chlorine. However, the criteria for the assignment of the molecular formulae and elemental compositions to prioritise appear overly subjective (e.g. Koch *et al.*, 2007; Cortés-Francisco *et al.*, 2014; Gonsior *et al.*,

2014; Herzsprung *et al.*, 2014; Mangal *et al.*, 2016; Hollender *et al.*, 2017). Furthermore, many micropollutants which affect the environment do not contain such heteroatoms, thereby restricting the ability of the approach to distinguish anthropogenic compounds from the natural derived background DOM (e.g. Gros *et al.*, 2006; Loos *et al.*, 2012; Santos *et al.*, 2013; Ruff *et al.*, 2015; Schymanski *et al.*, 2015).

Given the shortcomings of the above, a new approach is needed to exploit the potential offered by HRMS in the analysis of DOM. An alternative would be to use an untargeted approach, akin to that used in the "-omics" fields, wherein statistical methods are employed to define which compounds vary within a sample set (Marshall and Rodgers, 2004; Viant, 2008; Corilo *et al.*, 2010; Lankadurai *et al*., 2013). For example, in the field of environmental metabolomics untargeted analyses are used to explore a model organism's metabolic response to an external stressor, such as, drought, temperature or contaminant exposure (Lankadurai *et al*., 2013). Such an approach revolves around the testing of hypotheses that the environmental stressor will cause a change in the metabolome of an organism. The metabolome of "stressed" organisms determined by HRMS is compared to a control organism using multivariate statistics to investigate if the metabolome of the stressed organism is significantly different from that of a control. If the metabolomes are found to be significantly different, then a data reduction step is used to determine which metabolites vary between the control and stressed organism's metabolome. Once the discriminating compounds have been determined these metabolites are prioritised for identification. After identification these metabolites are related back to known metabolic processes to determine which aspect of the organism's metabolism has been affected by the external stressor. Critically, this approach relies on comparing complex MS datasets which can only realistically undertaken *in silico*. A similar approach could be applied to the untargeted analysis of organic micropollutants in surface water by comparing the compositions of different sources of DOM in the aquatic environment based on HRMS data.

As discussed in Chapter 1, multivariate statistics have been used to determine if there are significant compositional difference in DI-HRMS spectra produced by different extraction methods and from DOM originating from different environmental settings (e.g. Chen *et al.*, 2014; Dubinenkov *et al.*, 2015; Li and Minor, 2015; Li *et al.*, 2017). However, such methods have yet to be applied to point sources of anthropogenic pollution, where targeted approaches are most commonly used. As in previous studies of DOM, water could be sampled from around different point sources along a river, extracted and analysed using DI-HRMS. The DI-HRMS

spectra of the extracts could be compared using statistical methods to determine which DOM components differ between the waters sampled from different locations.

PCA is one of the most commonly used statistical methods for determining if qualitative differences exist in DI-HRMS spectra of DOM extracts (e.g. Chen *et al.*, 2014; Dubinenkov *et al.*, 2015; Li and Minor, 2015; Li *et al.*, 2017). However, significant improvements are required to improve the way PCA of DI-HRMS spectra of DOM extracts are used. For example, some previous studies, which use PCA to compare DI-HRMS spectra of DOM, only use single replicates of the extracts, thereby failing to account for any variations caused by the analysis or extraction of DOM from water (Chen *et al.*, 2014; Li and Minor, 2015; Li *et al.*, 2017). Furthermore, in "-omics" studies the use of quality controls (QCs) has been shown to be vitally important to assess errors in instrument performance and/or changes caused by pre-processing of the data. An appropriate QC should be a qualitative representation of all the samples being analysed and should be analysed repeatedly during the analytical sequences (Kirwan *et al.*, 2014).

While there are many different point sources of micropollutants into rivers sewage treatment works have been the focus of a number of targeted analysis studies, revealing discharges of a range of synthetic industrial compounds including, pharmaceuticals, flame retardants, surfactants and illicit drugs (Segev *et al.*, 2009; Karolak *et al.*, 2010; Loos *et al.*, 2012; Verlicchi *et al.*, 2012; Sheng *et al.*, 2014). Therefore, comparing the chemical composition of the background riverine and sewage outfall DOM using "-omics-based" statistical approaches might be a productive avenue of investigation.

3.2 Aims

The overall objective of this study is to describe the development of direct infusion Orbitrap high resolution mass spectrometry to compare the composition of riverine DOM to sewage effluent DOM using water collected from the River Chew and Chew Stoke sewage works outfall.

The specific aims of this chapter are to:

(i) Explore the nature of mass spectra of high-resolution DI-HRMS of riverine and sewage effluent DOM.

(ii) Investigate the repeatability of DI-HRMS analysis of DOM and its application to the assessing SPE extraction of DOM from river water

(iii) Determine whether qualitative differences exist in the chemical composition of DOM in water upstream and downstream of the sewage outfall.

(iv) Determine if the DOM extracts can be differentiated using multivariate statistical analysis of the high-resolution mass spectra.

(v) The application of univariate statistics will be used to determine which ions (tentatively = DOM molecular species) are most significant in discriminating between the different DOM extracts.

A sewage outfall was chosen because, as discussed previously, sewage outfalls are very significant contributors of DOM to rivers but have only ever been assessed using targeted analysis. However, it is hypothesised that *processed sewage effluent will be a highly complex mixture of compounds many of which will be overlooked in a targeted analysis study, but still may be of environmental relevance*. Only by undertaking comprehensive untargeted analyses can such compounds be identified and their environmental load and ecotoxicity be determined.

3.3 Site information

The Chew Stoke sewage treatment works situated on the River Chew, Somerset, UK, downstream of Chew Valley Lake, was selected as the site to test and develop this method. Chew Valley Lake is a drinking water reservoir which was constructed in the 1950's, shown in Figure 3.1. Water from the reservoir is used to regulate the flow in the River Chew to sustain the local ecosystem. The treatment methods used at Chew Stoke sewage treatment works are; (i) preliminary solid removal to remove large particulates, (ii) primary settling to further allow smaller particulates to flocculate and settle out of the water column, (iii) secondary treatment using trickle filter beds to biologically break down the organic matter. The sewage treatment works has a tertiary treatment which includes phosphorous stripping. The final treated effluent is discharged into the river downstream of the reservoir.

The surrounding landscape has a clay soil type. The river was assessed visually for additional point sources, such as discharge pipes and drainage ditches between the upstream and downstream sampling sites to mitigate other factors which could cause a change in the DOM composition. The distance from the lake spillway to the sewage outfall is approximately 650 m and there are no dwellings on the banks of the river in the immediate catchment which could have septic tanks discharging into the river. There are low levels of cattle grazing on the fields surrounding the sewage treatment works.

Figure 3.1 shows the sampling site 62 m upstream of the sewage treatment outfall from which the upstream water was collected. The sewage outfall water was collected directly from the discharge outfall culvert. The downstream water sample was collected 48 m downstream of the sewage outfall.

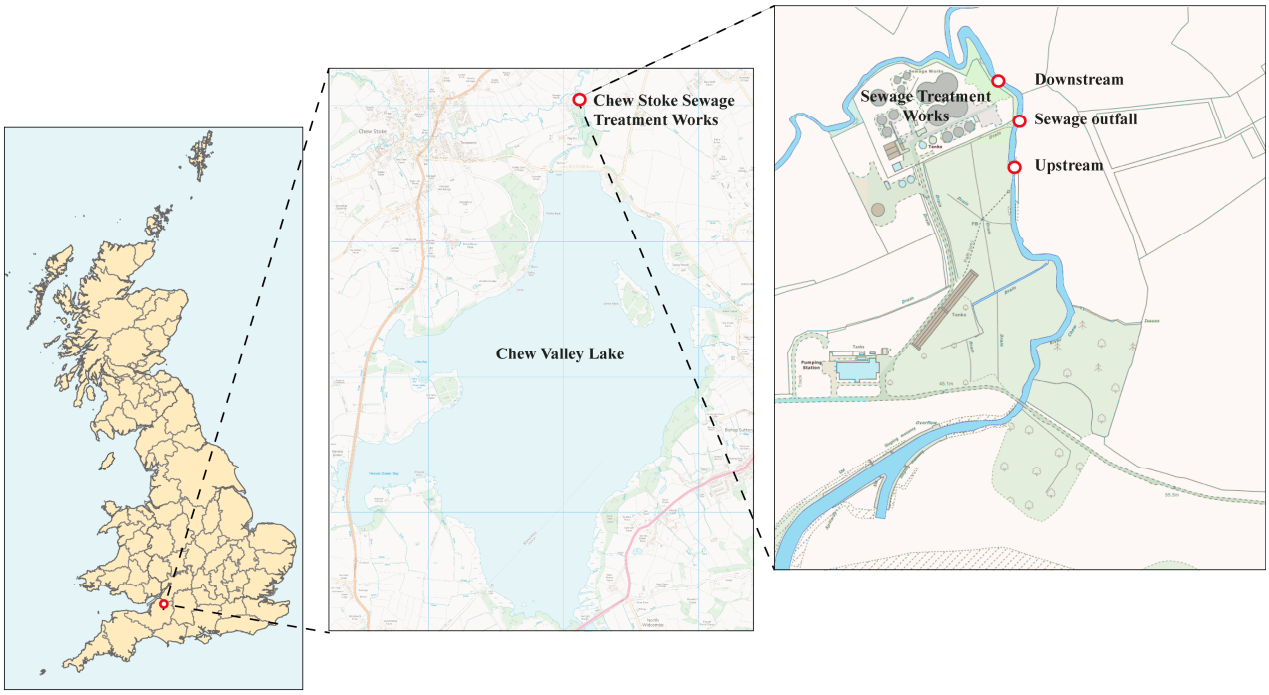


Figure 3.1. A map of the UK with Chew Stoke sewage treatment works marked (o), Chew Valley Lake and River Chew north of the reservoir and Chew Stoke sewage treatment works (o). A map of the sampling sites marked and labelled where the water was collected for this study.

The river's primary source of water is the lake. Therefore, the river DOM composition and flow is well controlled in comparison to other river systems, where the river may be an admixture of tributaries and point sources. Lake DOM has been shown to have a generally temporally consistent composition. Kellerman *et al.*, (2014) compared the composition of 120 lakes across Sweden. This study concluded that the chemodiversity of DOM when residing in a lake over time decreases due to in-lake processes such as mixing, photodegradation, sedimentation and biodegradation. The long residence time of water in lakes allows equilibration/homogenisation by these various degradation processes. Furthermore, changes in lake DOM composition due to episodic, hydrological or anthropogenic inputs, are small in comparison to the size of the lake and are therefore heavily diluted, making the DOM composition more consistent.

3.4 Analysis of DOM from Chew Stoke

There was no visual difference between the water samples. The water samples and comparative HPLC water were filtered and extracted as described in Section 2.2, within 48 h of collection. The dry extracts were dissolved in methanol/water (1:1 *v/v*, 1 ml). DOM extracts of the water collected upstream and downstream were yellow coloured and clear, the extracts from the sewage outfall was noticeably darker than the riverine DOM extracts. The blank extracts were clear and colourless. The volume of water extracted, the extraction method and the volume of solvent used to dissolve the SPE extracts was consistent for all samples in order allow the direct comparison of the composition of the DOM in each extract.

3.4.1 DOC analysis of water samples and SPE extracts from Chew Stoke

DOC analysis was carried out as described in Section 2.3.1 to determine the concentration of organic carbon extracted. Table 3.1 shows the concentration of organic carbon in the filtered water collected from each site, the amount of organic carbon SPE extracted from the filtered water and the extraction efficiency of the SPE. There were no discernible differences between the concentrations of organic carbon in the filtered water samples from each site. The SPE extracted a similar concentration of organic carbon from each of the water samples. The extraction efficiencies of the SPE extracts were comparable to those achieved in previous studies which used HLB SPE cartridges to extract water (Raeke *et al.*, 2016; Li *et al.*, 2017).

The HLB cartridge is a cross linked styrene-based polymer solid phase with a hydrophilic group for increased retention of polar compounds. The extraction efficiencies show that not all organic compounds are extracted using the HLB SPE cartridge. The reasons for this could be because a compound does not bind to the solid phase and therefore pass through the cartridge or that compounds bind strongly to the solid phase and cannot be eluted using methanol. Therefore, only a fraction of the DOM within both the riverine and sewage effluent will be analysed. However, all extractions are carried out under the same conditions. Therefore, comparable compositional extracts will have been collected. Overall it is not possible to determine which compounds have been retained over those lost using DI-HRMS as the concentration of individual compounds is too low to be measured directly from the unextracted water sample. However, it may be possible using Fourier transform infrared spectroscopy, fluorescence or pyrolysis gas chromatography mass spectrometry to analyse the unextracted

and extracted water samples and determine the classes and functional groups which these cartridges extract.This would further the composition of the extracts being analysed,

Table 3.1. DOC of the filtered water, Amount of organic carbon extracted using SPE and the extraction efficiency of the extraction procedure

Samples	DOC filtered water (mgl^{-1})	DOC extracts (mg)	Extract efficiencies (%)
Upstream	3.39	1.50 ± 0.04	42.11 ± 0.5
Sewage Outfall	3.19	1.29 ± 0.05	40.49 ± 1.7
Downstream	3.45	1.40 ± 0.06	40.34 ± 0.8
Blank	N/A	0.18 ± 0.02	N/A

3.4.2 DI-HRMS of SPE extracts collected from Chew Stoke Sewage Treatment Works

All of the replicate SPE extracts from the sampling sites and the blank SPE extracts were analysed by DI-HRMS in positive ion mode as described in Section 2.3.2. Previously reported DI-HRMS spectra of extracted DOM have been shown to comprise of an extremely complex distributions of ions (Flerus *et al.*, 2011; Osborne *et al.*, 2013; Mangal *et al.*, 2016); similarly complex spectra were obtained here. Figure 3.2(a) shows the DI-HRMS spectrum of an upstream SPE water extract for the *m/z* 150 to 2000. This spectrum is similar to those previously reported in natural environments (Dalzell *et al*., 2009; Hockaday *et al.*, 2009; Osborne *et al.*, 2013). It shows an overall normal distribution of ion intensities in the mass range of *m/z* 150 to 800. Figure 3.2(b) shows ions within the 50 Da mass range from *m/z* 300 to 350. This 50 Da window shows that there are a cluster of peaks at 1 Da intervals. This regular 1 Da spacing is the result of the majority of atomic mass being comprised of nucleons for which both protons and neutrons have a nominal mass of 1.

Figure 3.2(c) shows a 3 Da mass window of the upstream DI-HRMS spectrum which shows the individual ions of three of the 1 Da spaced clusters seen in the 50 Da mass range (Figure 3.2(b)). Each of the clusters show between 9 and 12 individually resolved isobaric ions. This illustrates the exceptional complexity of DOM, as each molecular ion could potentially be a unique compound. However, many of these ions will be different adducts or isotopes of the same compound. This emphasises the difficulty of comparing these spectra with those reported previously, because, while the overall DI-HRMS ion envelop may appear to possess a similar form for different extracts, the individual ions making up the spectra may differ substantially.

At this stage these raw mass spectra provide no insights into the nature of the DOM represented by the spectra.

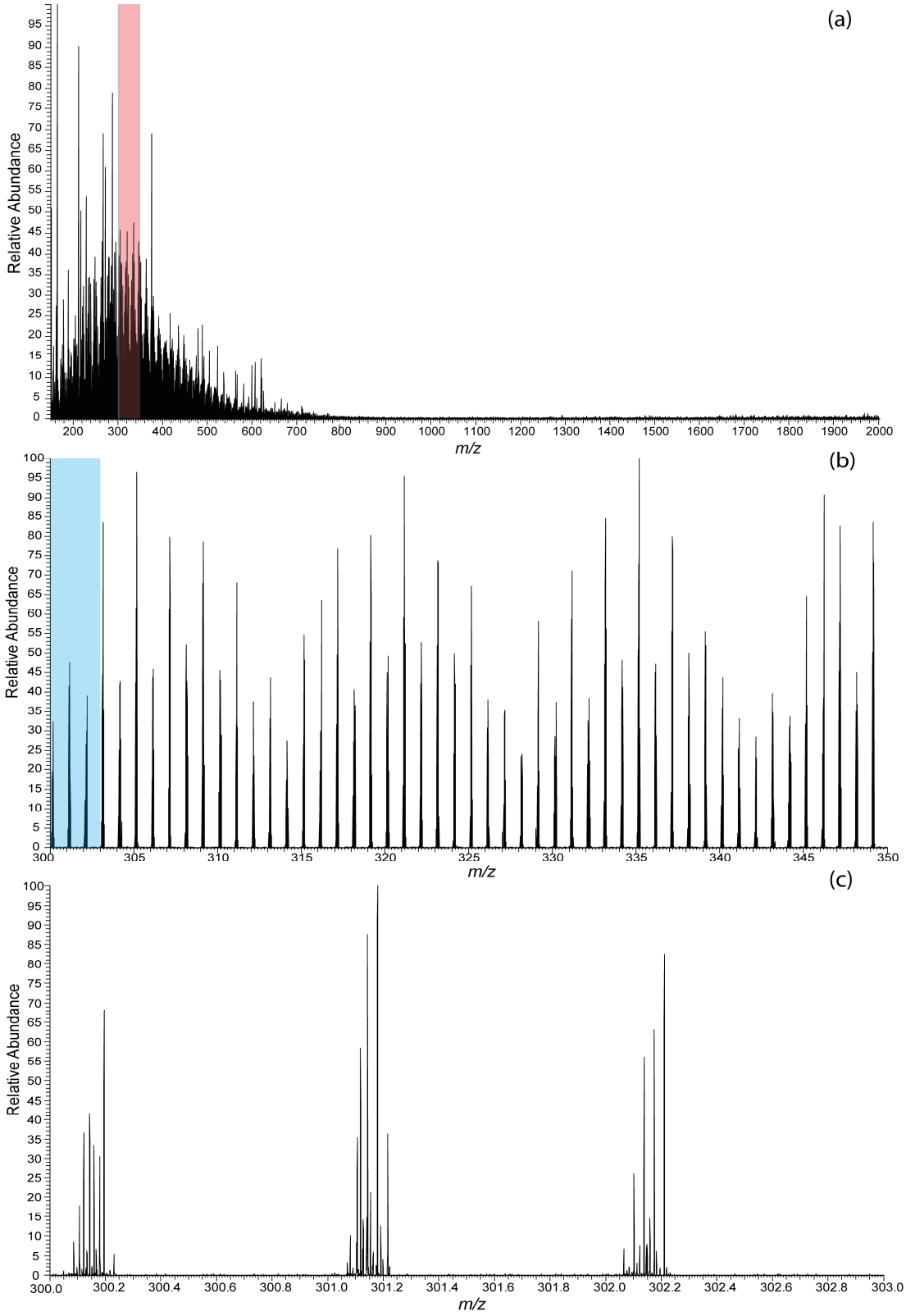


Figure 3.2. A representative DI-HRMS spectrum of an upstream SPE extract showing 3 different mass ranges: (a) full range *m/z* 150 to 2000 with the 50 Da range highlighted in red and 3Da range highlighted in blue, (b) 50 Da range *m/z* 300 to 350 with the 3 Da mass range highlighted in blue (c) 3 Da range *m/z* 300 to 303.

3.4.2.1 Comparison of DI-HRMS spectra of the replicate SPE extracts of each water sample

Before comparing the DI-HRMS spectra of different SPE extracts from each sampling site, it is important to understand the reproducibility of these complex DOM spectra and the variance between replicate extracts and analytical blanks.

Figure 3.3 shows the DI-HRMS spectra of the extracted HPLC grade water. The high sensitivity of the Orbitrap means that even unextracted HPLC grade water yields a reasonably complex spectrum. While many of the ions relate to solvent clusters, there is clearly a consistent contamination derived from the cartridge and/or concentrated from the HPLC water. Importantly, the DOC analysis (Table 3.1) showed that the concentration of carbonaceous material in the blank is much lower than the other SPE extracts of DOM. Figure 3.4 compares two narrower mass windows around the ions *m/z* 183.9496 and *m/z* 268.0025 which have the highest abundance seen in the DI-HRMS spectra of the blank extracts to the DI-HRMS spectrum of the upstream extract. The ions observed in the blank extract are undetectable the upstream DI-HRMS spectra of the SPE extract. This is likely due to the difference in concentration of the compounds found in the upstream in comparison to those in the blank extract as shown by the TOC, therefore, the signal of ions seen in the blank is supressed. Critically, the ions seen in the upstream extracts are not present in the blank, showing the presence of compounds specific to the river DOM.

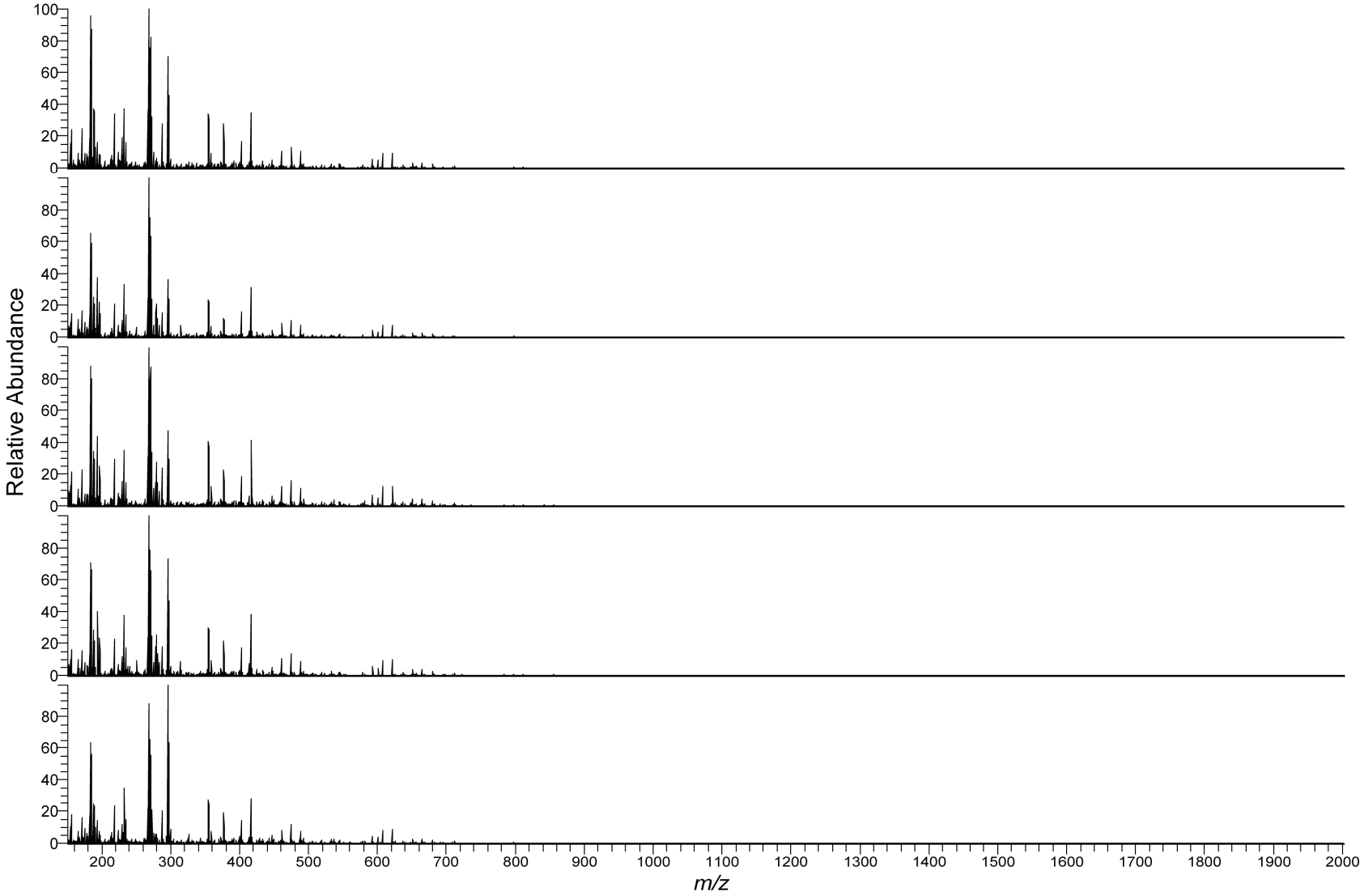


Figure 3.3. Comparison of the DI-HRMS spectra of the five replicate blank HPLC water SPE extracts *m/z* 150 to 2000.

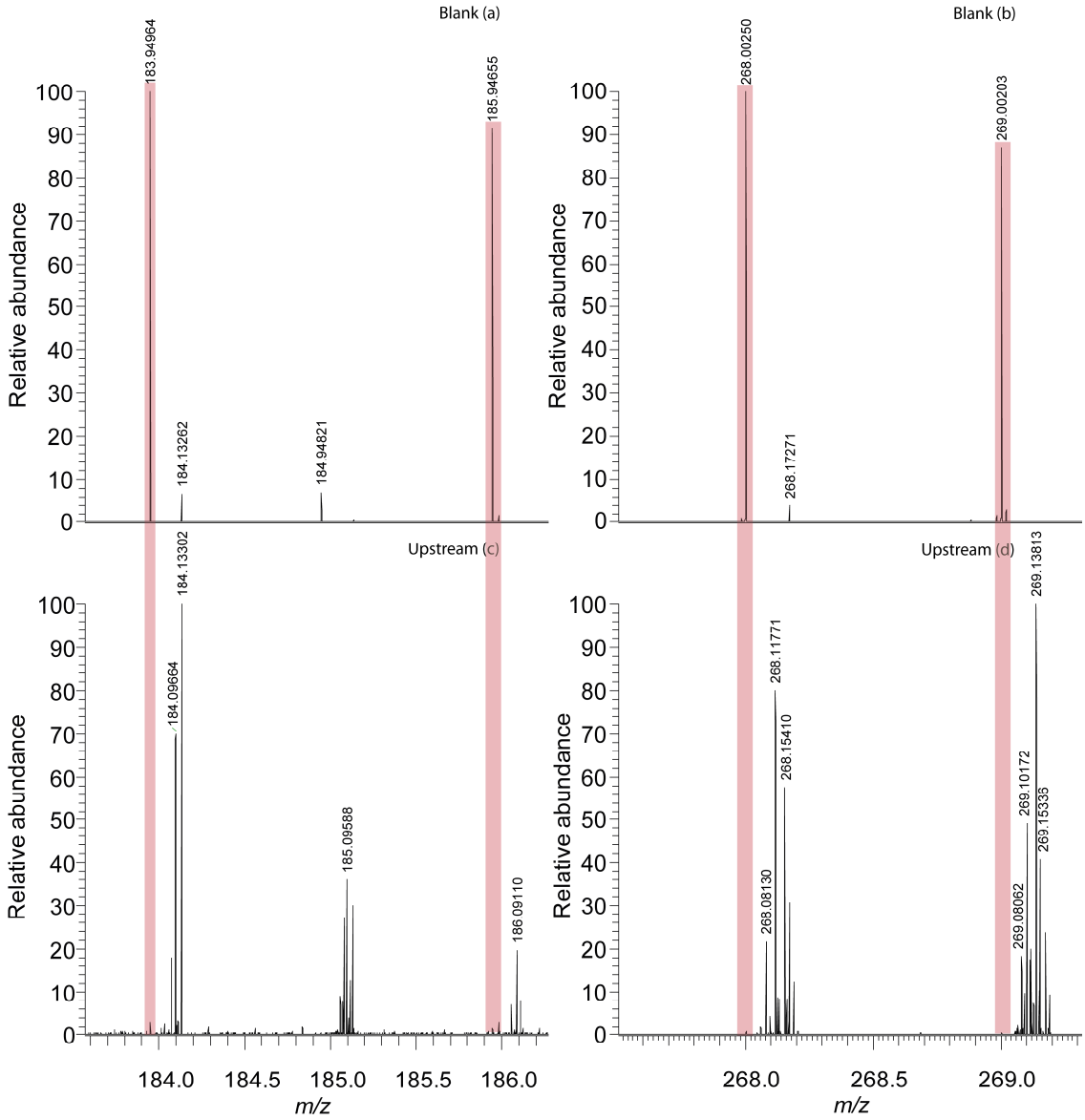


Figure 3.4. Comparison of selected regions of the DI-HRMS spectra displaying highest abundance ions in the blank SPE extract to the upstream extract: (a) Blank extract DI-HRMS mass spectrum in the range *m/z* 183.55 to 186.25, (b) Blank DI-HRMS mass spectrum in the range *m/z* 267.50 to 269.31, (c) Upstream extract DI-HRMS spectrum in the range *m/z* 183.55 to 186.25 (d) Upstream extract DI-HRMS spectrum in the mass range *m/z* 267.50 to 269.31. Red boxes highlight the *m/z* of ions with the highest intensity ions in the blank DI-HRMS spectra on the axis of the upstream extracts DI-HRMS spectra.

Figure 3.5 shows a comparison of DI-HRMS spectra of the upstream replicate SPE extracts. There is clearly a normal distribution of abundances of ions spanning *m/z* 150 to 800 mass range common to all DI-HRMS spectra of the upstream extracts. However, there are clearly 8 ions (highlighted in the second spectrum using *) between *m/z* 490 to 805 (highlighted in red) which are visible in some extracts but not in others. These 8 ions are not observed in any of the blank extracts DI-HRMS spectra and therefore are not coming from the extraction procedure. In comparison to the blank extracts DI-HRMS spectra shown in Figure 3.3 there is a clear difference in the overall composition of the upstream extracts.

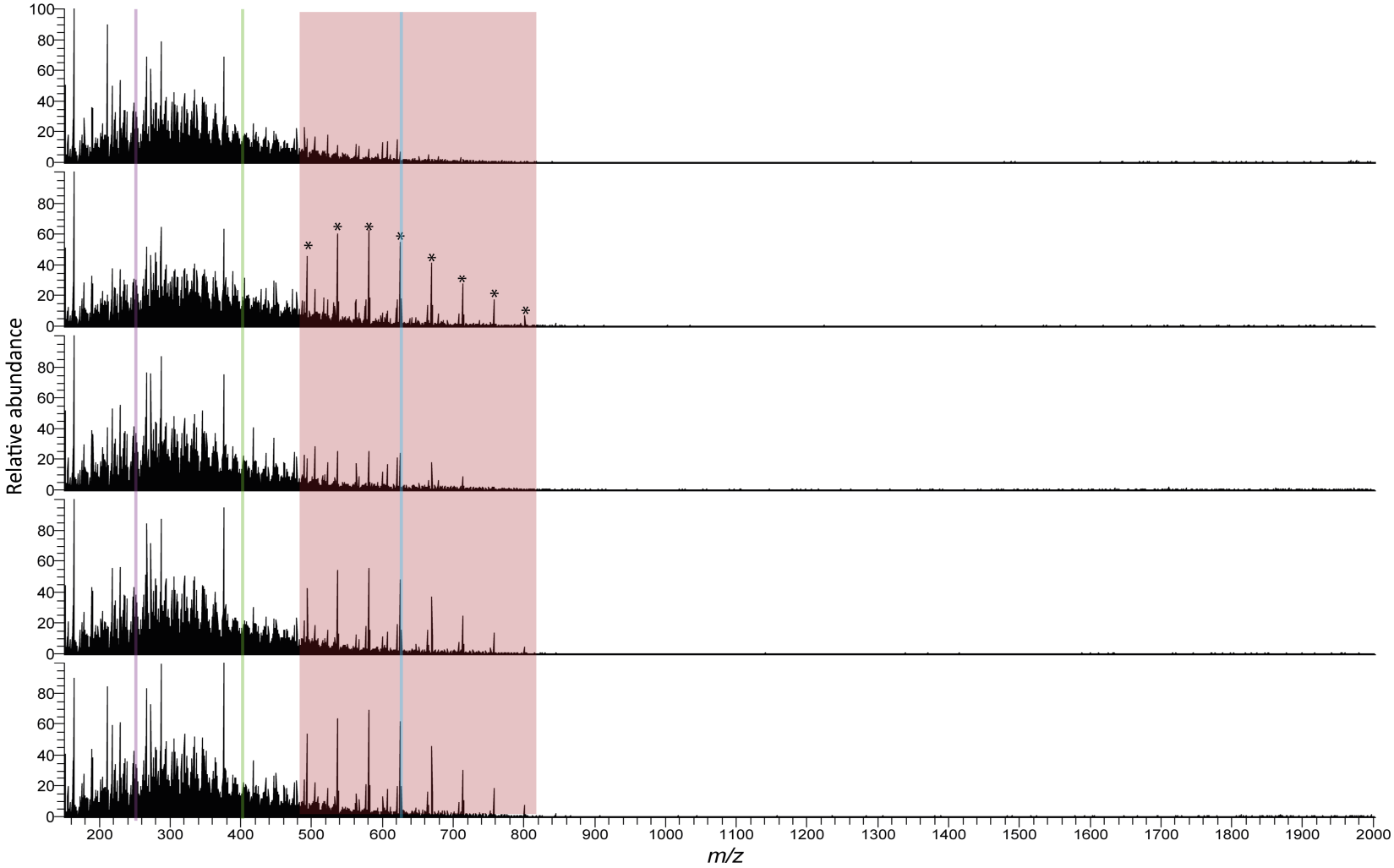


Figure 3.5. Comparison of DI-HRMS spectra of the five replicate SPE extracts of upstream water collected from Chew Stoke sewage treatment works displayed over the range *m/z* 150 to 2000. Red box highlights the mass spectral range which contains the eight ions each indicated with a * which are changing in abundance when comparing the upstream DI-HRMS spectra. The purple, green and blue boxes highlight the mass ranges, *m/z* 250 to 255, *m/z* 400 to 405, and *m/z* 625 to 630 which are used in Figure 3.6.

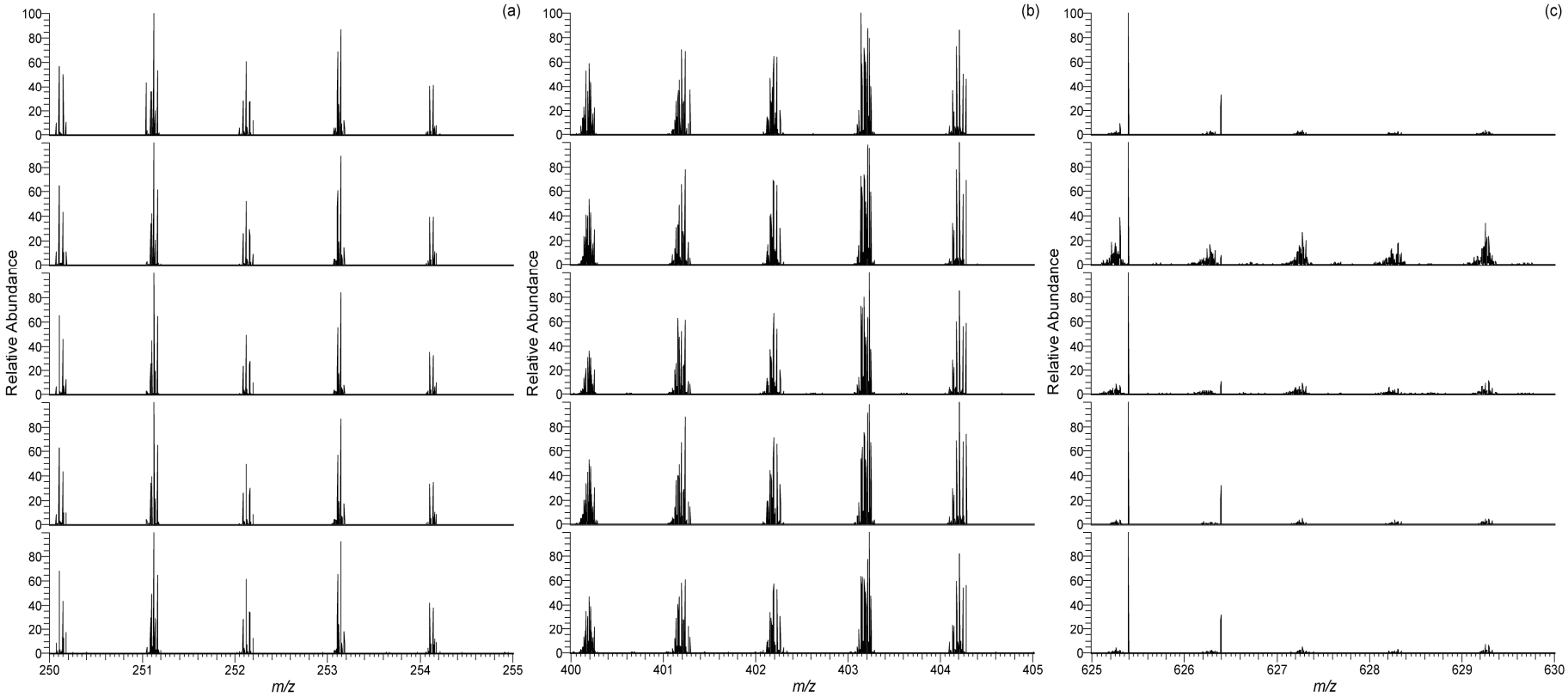


Figure 3.6. Comparison narrow mass ranges of the upstream DI-HRMS spectra for the 5 SPE extract replicates for the mass ranges (a) *m/z* 250 to 255, (b) *m/z* 400 to 405, and (c) *m/z* 625 to 630.

Figure 3.6 compares 3 narrower mass ranges of the DI-HRMS spectra of the upstream SPE replicate extracts to explore variations in the individual ion distributions and abundances. All ions in each of the 3 mass ranges selected are present in all 5 replicate extracts showing reproducibility in the extraction and analysis protocol. For the most part the distributions of ions between the replicates show remarkable consistency. However, as highlighted in Figure 3.5 there are ions which vary in abundance between the extracts, e.g. *m/z* 625.391 is present in all the replicates despite it not being visible in the full mass range DI-HRMS spectra. Reassuringly, these results show that the qualitative compositions are quite consistent even if there are changes in individual ion abundances. This could be due to a difference in variation of these compounds' extraction efficiency or analytical variability.

Figure 3.7 compares the DI-HRMS spectra of the sewage outfall replicate SPE extracts for the mass range *m/z* 150 to 2000. The DI-HRMS spectra are remarkably similar for all the replicates. Interestingly, all replicates display series of ions with a regular mass difference, seemingly indicative of the presence of an homologous series of compounds. These spectra are very different to those obtained for the upstream extracts and those previously for DOM from natural lake waters (Flerus *et al.*, 2011; Osborne *et al.*, 2013), suggesting significant anthropogenic contributions from the sewage outfall.

Figure 3.8 compares narrower mass ranges of the DI-HRMS spectra of the sewage outfall SPE replicates. These mass ranges correspond to the same ranges used in comparing the DI-HRMS spectra of upstream SPE replicates above, thereby allowing comparisons to be made between the sewage outfall and upstream DI-HRMS spectra. Figure 3.8 shows that for all three of the narrower mass ranges of the sewage outfall DI-HRMS spectra, all ions are seen in all replicates. Furthermore, the ion abundances are similar across all spectra showing that the analyses and extractions were highly repeatable. Figure 3.8 shows the spectra are characterised by clusters of ions with 1 Da spacings, akin to the pattern observed in the spectra of the upstream extracts. However, whereas the clusters of ions in the upstream DI-HRMS spectra exhibited generally similar abundances relative to the other ions present, the sewage outfall DI-HRMS spectra contained several ions of much more variable and higher abundances.

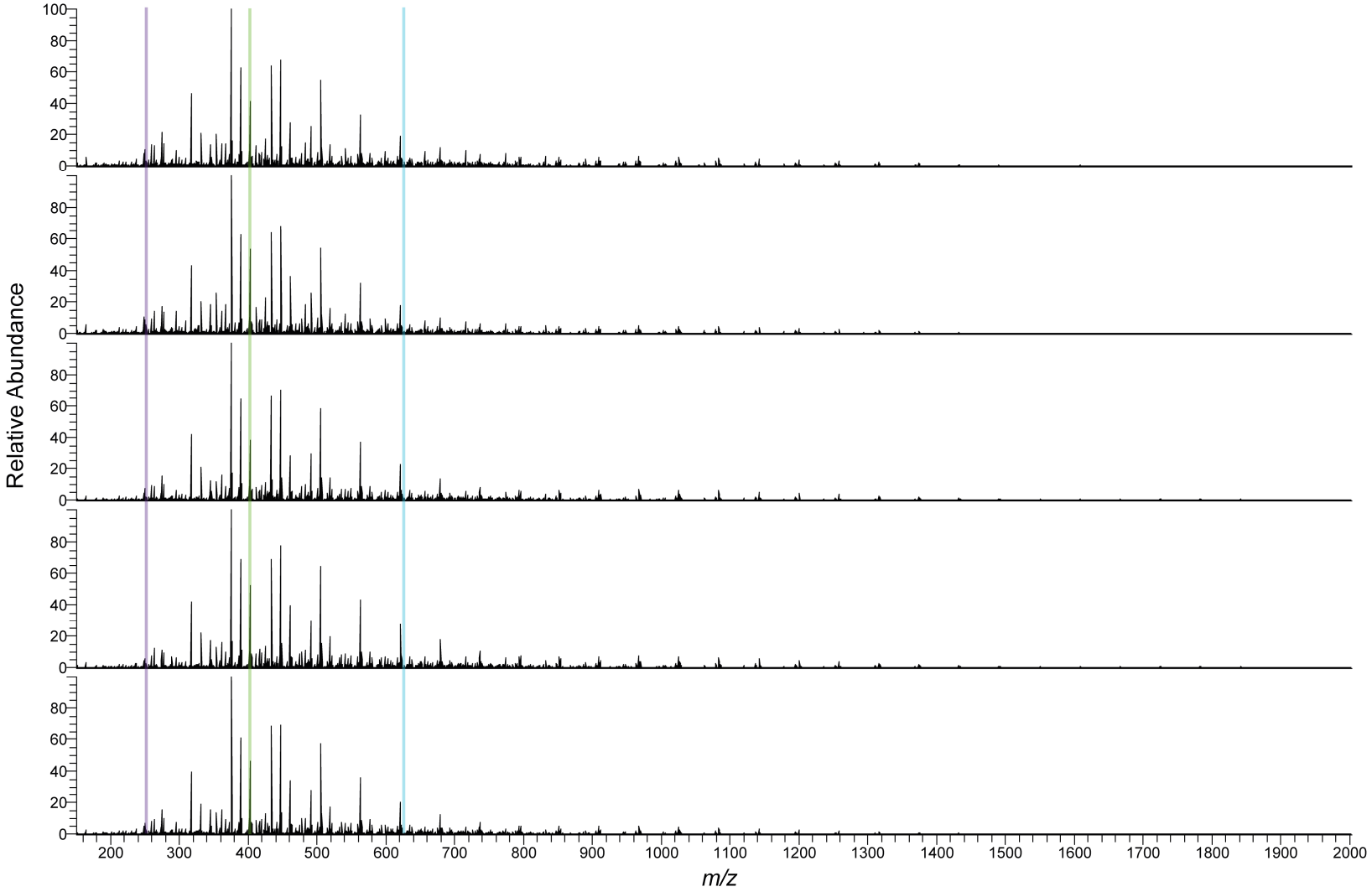


Figure 3.7. Comparison of DI-HRMS spectra of the five replicate SPE of sewage work discharge waters collected from Chew Stoke sewage treatment works for range *m/z* 150 to 2000. The purple, green and blue boxes highlight the mass ranges, *m/z* 250 to 255, *m/z* 400 to 405, and *m/z* 625 to 630 which are used in Figure 3.8.

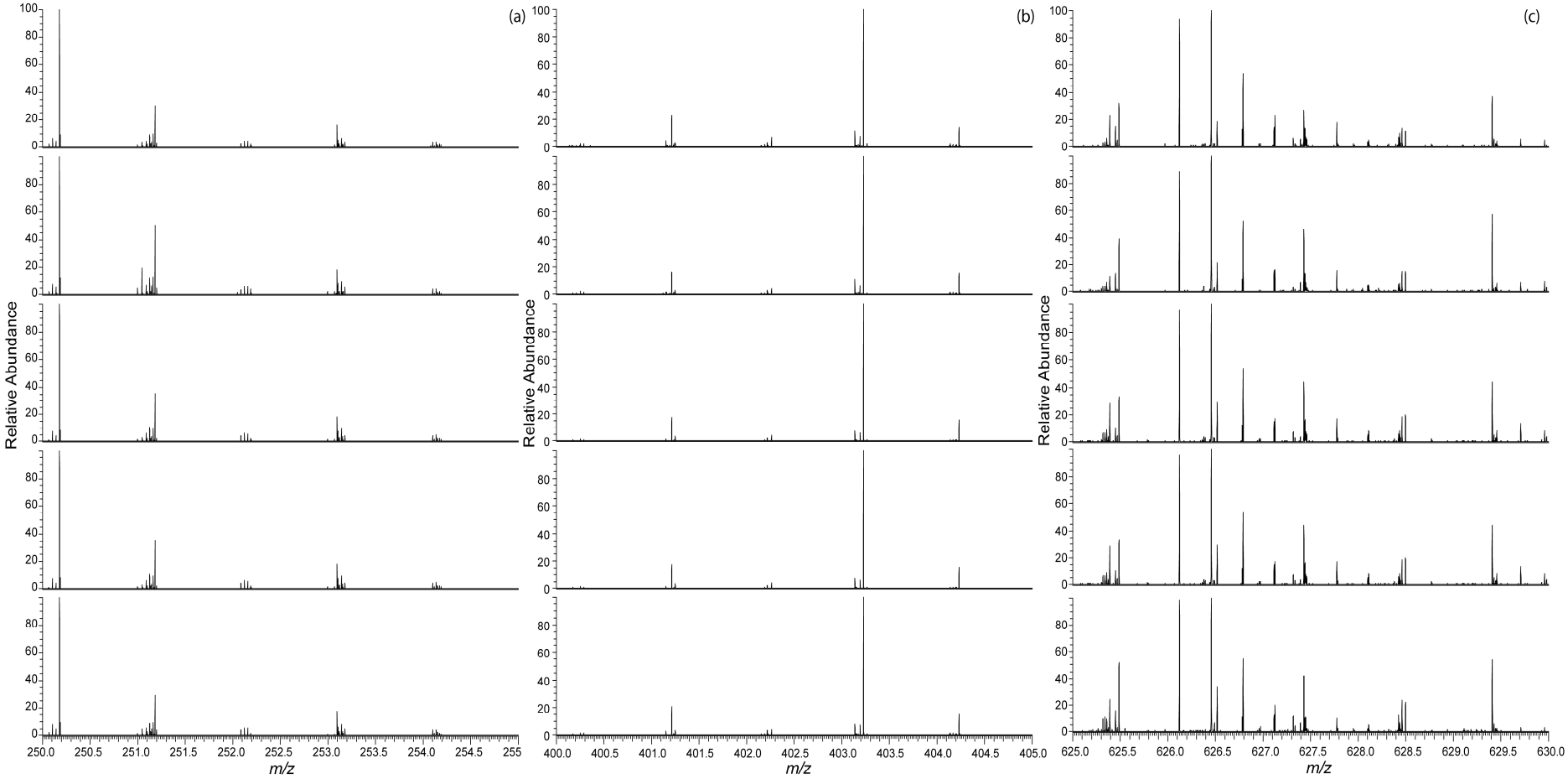


Figure 3.8. Comparison of narrower mass ranges of the sewage outfall DI-HRMS spectra of the 5 SPE extract replicates display for the mass ranges: (a) *m/z* 250 to 255, (b) *m/z* 400 to 405, and (c) *m/z* 625 to 630.

Figure 3.9 shows the DI-HRMS spectra of the downstream SPE replicate extractions for the mass range *m/z* 150 to 2000. Overall, the composition of the downstream DI-HRMS spectra is similar across all the replicates. However, the ion at *m/z* 212.118 shows a clear variation in abundance across the different DI-HRMS spectra of the downstream replicate extracts. Compared to the upstream and sewage outfall DI-HRMS spectra the downstream DI-HRMS spectra are more similar in composition to the sewage outfall DI-HRMS spectra, reflecting the contribution of the sewage outfall to the overall river DOM composition. Both the sewage outfall and downstream DI-HRMS spectra contain ions with the same a regular mass difference, indicative of the presence of an homologous series of compounds.

Figure 3.10 shows three narrower mass ranges for the DI-HRMS spectra of the downstream replicate extractions to compare individual ions with the same mass ranges displayed as used in Figure 3.6 and Figure 3.8, thereby allowing comparison of the downstream, upstream and sewage outfall DOM extracts. The DI-HRMS spectra of the downstream extraction replicates show all ions in the *m/z* 250 to 255 and *m/z* 400 to 405 mass ranges are present and at similar abundances. However, Figure 3.10(c) displays the mass range *m/z* 625 to 630 and in one replicate *m/z* 625.391 appears less abundant compared to other four DI-HRMS spectra of the downstream replicates.

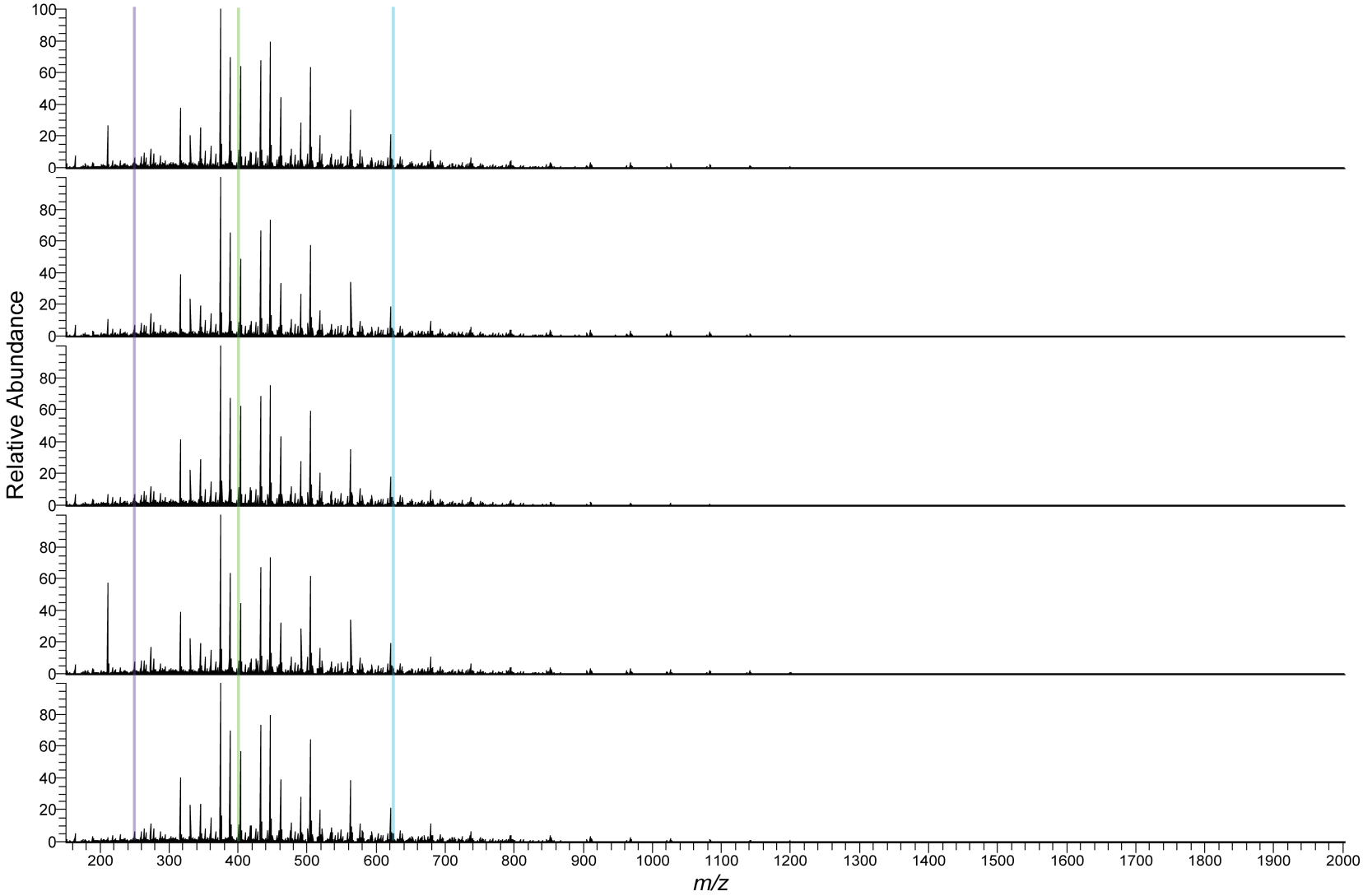


Figure 3.9. Comparison of DI-HRMS spectra of the five replicate SPE of downstream water collected from Chew Stoke sewage treatment works displaying the range *m/z* 150 to 2000. The purple, green and blue boxes highlight the mass ranges, *m/z* 250 to 255, *m/z* 400 to 405, and *m/z* 625 to 630 which are used in Figure 3.10.

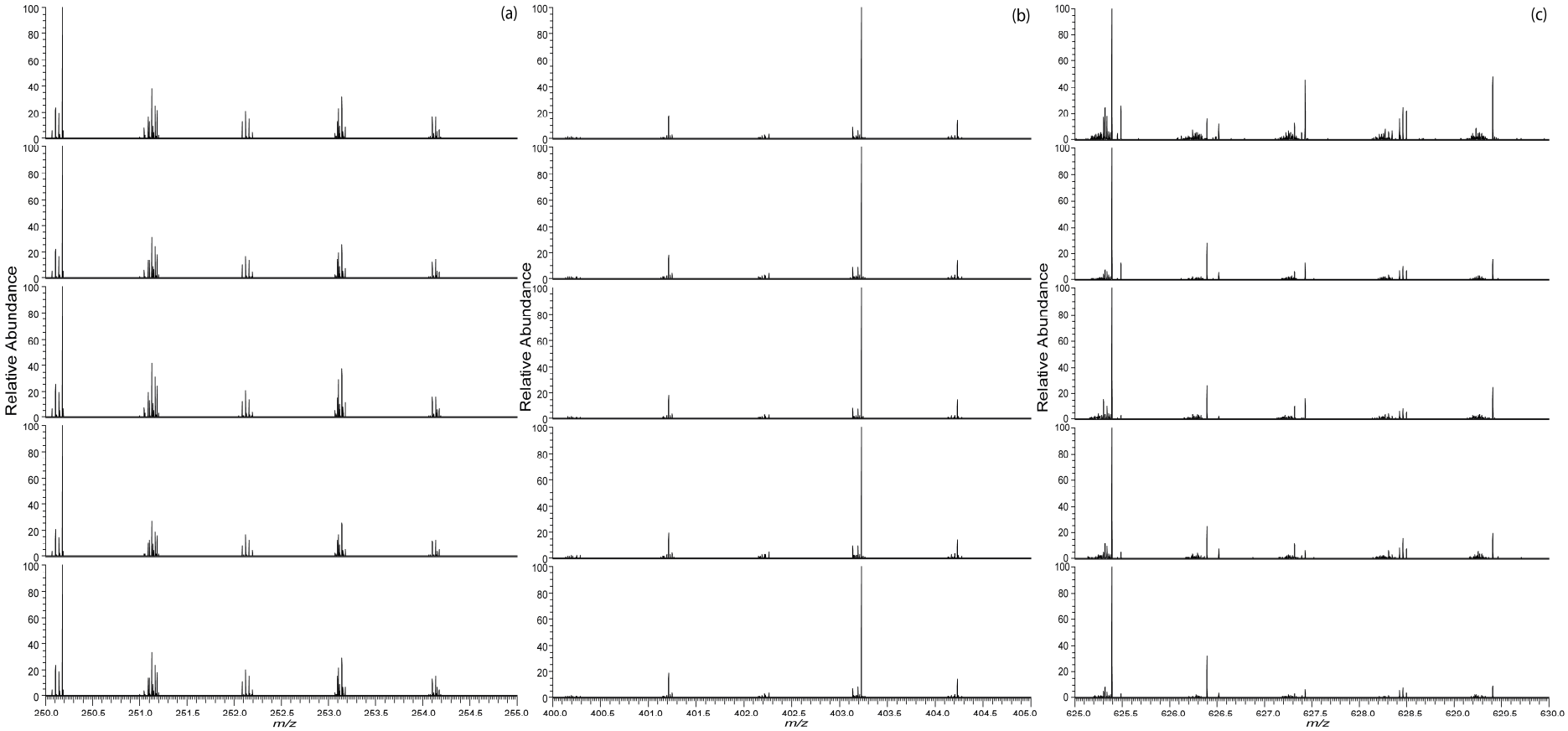


Figure 3.10. Comparison of selected narrow mass ranges of the downstream DI-HRMS spectra of the 5 SPE extraction replicates across the mass ranges: (a) *m/z* 250 to 255, (b) *m/z* 400 to 405, and (c) *m/z* 625 to 630.

Comparing individual ions in the narrower selected mass ranges of the DI-HRMS spectra of the downstream, sewage outfall and upstream extract (Figure 3.6, Figure 3.8 and Figure 3.10) reveals clear similarities and differences across the three sampling sites, which are consistent across all replicates. As shown in Figure 3.6(a), Figure 3.8(a) and Figure 3.10(a) in the mass range *m/z* 250 to 255 shows that the ions present in the upstream DI-HRMS spectra are also seen in the downstream and sewage outfall DI-HRMS spectra. However, there are additional ions present in the downstream and sewage outfall DI-HRMS spectra not seen in the upstream spectra, the most prominent being *m/z* 250.180 and *m/z* 251.185. The ions only present in the downstream and sewage outfall extract spectra have a higher abundance in the sewage outfall DI-HRMS spectra compared to the downstream DI-HRMS spectra, which clearly relates to dilution of the sewage discharge by the river.

Comparing the DI-HRMS spectra in Figure 3.6(b), Figure 3.8(b) and Figure 3.10(b) for the mass range *m/z* 400 to 405, the composition of the downstream and sewage outfall are clearly different to the DI-HRMS spectra of the upstream extracts. The upstream DI-HRMS spectra show clusters of multiple isobaric ions at 1 Da intervals, in comparison the downstream and sewage outfall DI-HRMS spectra contain several discrete ion clusters, the most intense ions being *m/z* 401.214, *m/z* 403.229 and *m/z* 404.233, which are not found in the upstream DI-HRMS spectra. However, the ions in the upstream DI-HRMS spectra are found in the sewage outfall and downstream DI-HRMS spectra but at lower abundance compared to the additional ions in the sewage outfall and downstream extract DI-HRMS spectra.

Comparing the DI-HRMS spectra of the different SPE extracts of the water collected from the three sampling sites shown in Figure 3.6(c), Figure 3.8(c) and Figure 3.10(c) for the mass range *m/z* 625 to 630 shows that the DI-HRMS spectra for the upstream, downstream and sewage outfall extracts all differ in composition. The DI-HRMS spectra of the upstream replicates, as discussed above, show clusters of 1 Da spaced isobaric ions and the ion *m/z* 625.391 varies in abundance across all the extraction replicates. This ion is also seen in the sewage outfall and downstream DI-HRMS spectra. The downstream DI-HRMS spectra contain additional ions, which are not present in the upstream extract DI-HRMS spectra but are present in the sewage outfall DI-HRMS spectra. The most abundant ions are *m/z* 626.457, *m/z* 627.318 and *m/z* 629.409. However, the sewage outfall spectra contain unique ions not seen in either the upstream or downstream DI-HRMS spectra, notably *m/z* 626.123, *m/z* 626.791, *m/z* 627.126. This could be the result of the compounds coming from the sewage outfall being: (i) diluted in the river below the limit of detection, (ii) processed instream either through biodegradation,

and/or (iii) becoming bound to particulate matter which is removed by filtration during sample preparation.

Figure 3.11 shows the DI-HRMS spectra of the mixed QC solutions for the range *m/z* 150 – 2000, which were recorded after every five DI-HRMS analyses. The QC DI-HRMS spectra are in the figure ordered from the start of the sequence, 5, 10 and 15 analyses and the final analysis, ordered from top to bottom. Comparing the DI-HRMS spectra of the pooled QC solution shows no detectable differences in composition or ions abundances observed throughout the analytical sequence, confirming little or no variation in instrument performance throughout the analyses of the DI-HRMS of DOM extracts. The DI-HRMS spectra of the QC most closely resembles the composition to the downstream and sewage outfall extracts, which was entirely expected as the QC solution is a mixture of all the SPE extracts, and therefore should be similar in composition to the three DOM extracts.

Figure 3.12 compares three narrower mass ranges of the DI-HRMS spectra of the pooled QC standard. The DI-HRMS spectra shown in Figure 3.12 (a) and (b) of the mass ranges *m/z* 250 to 255 and *m/z* 400 to 405 the spectra are identical in every respect. However, Figure 3.12(c) shows variations ion abundances within the mass range *m/z* 625 to 630, however, these are much smaller than the changes observed between DI-HRMS spectra of the DOM SPE extracts and replicate extractions. While the reason for these differences are unknown, they are minor and will be taken account of in subsequent statistical analyses.

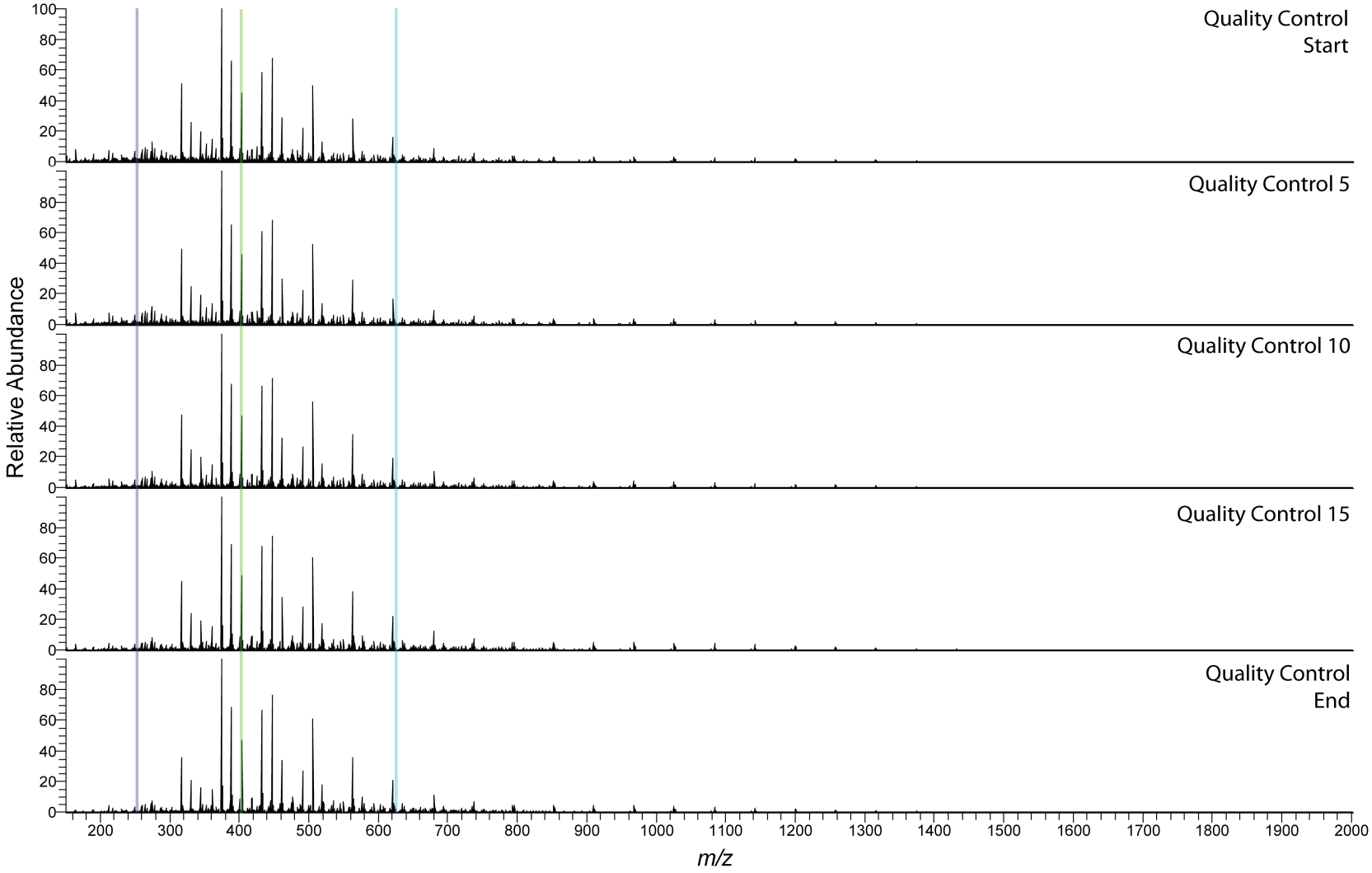


Figure 3.11. Comparison of 5 replicate DI-HRMS spectra of the pooled QC solution displayed across the range *m/z* 150 to 2000. The QC DI-HRMS spectra are ordered from start to end of the DI-HRMS run sequence: QC start of the analytical analyses (top), after 5, 10, 15 analyses and at the end of the analysis (bottom). The purple, green and blue boxes highlight the mass ranges, *m/z* 250 to 255, *m/z* 400 to 405, and *m/z* 625 to 630 which are used in Figure 3.12.

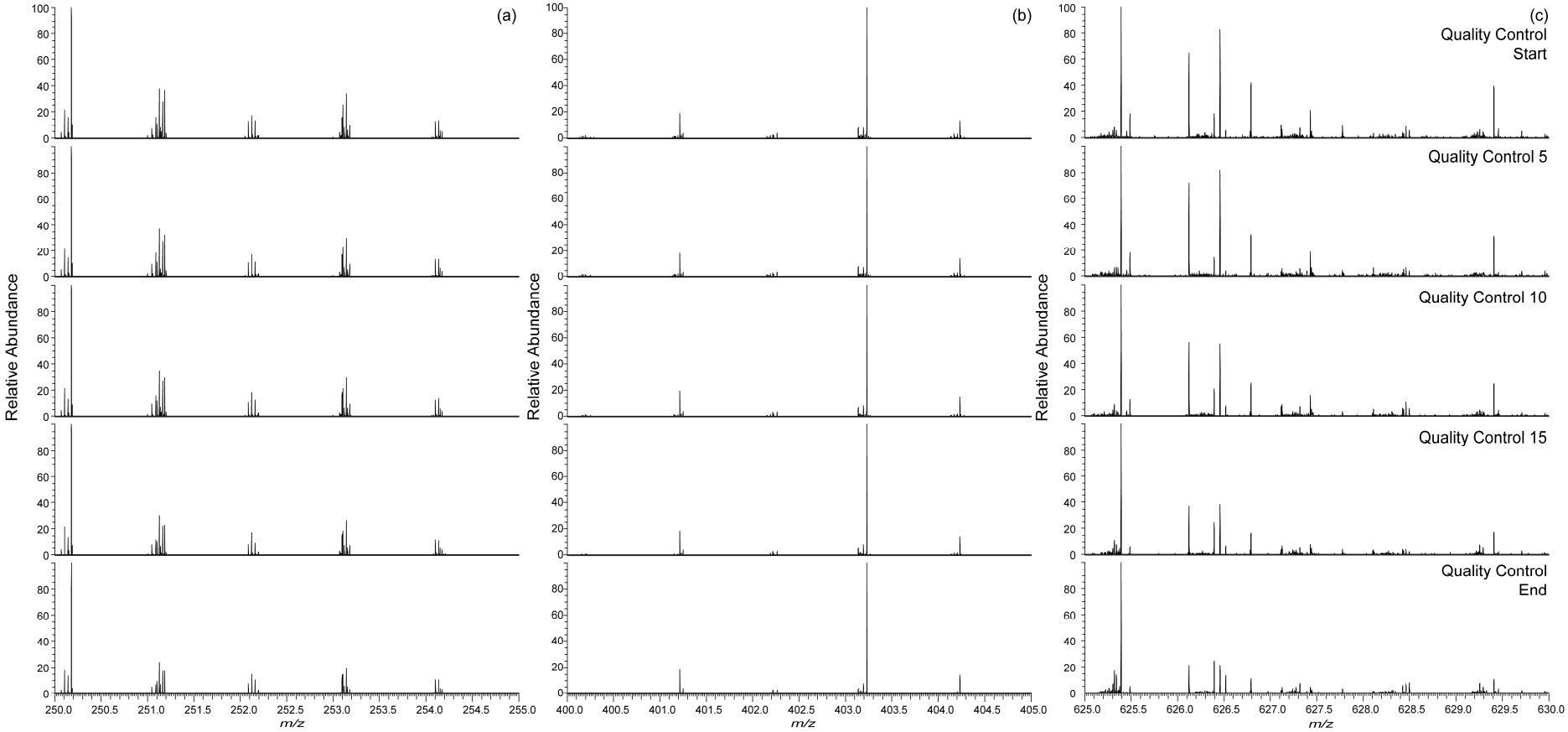


Figure 3.12. Comparison of selected narrower mass ranges of the pooled QC DI-HRMS spectra of the five replicate analyses across the mass ranges: (a) *m/z* 250 to255, (b) mass range *m/z* 400 to 405, (c) mass range *m/z* 625 to 630. The QC spectra shown are ordered from start to end of the DI-HRMS analyses. First analysis (top), after 5, 10, 15 analyses and at the end of the DI-HRMS analyses (bottom).

3.4.2.2 Comparison of DI-HRMS spectra of upstream, downstream, sewage outfall and blank DOM SPE extract and pooled QC

Comparing the DI-HRMS spectra of replicate SPE extractions of the blank, upstream, downstream and sewage outfall water showed generally consistent compositions of replicate extractions for each sample group. Variation existed within each group of replicate DI-HRMS spectra. Some comparisons have been made between the DI-HRMS spectra of different sample groups. However, this section will compare specific regions of the DI-HRMS spectra which show differences and similarities between the SPE extracts of the upstream, downstream, sewage outfall, blank and the pooled QC.

Figure 3.13 shows representative DI-HRMS spectra for the mass range *m/z* 150 to 2000 of a DOM extract comparing the blank, downstream, upstream, sewage outfall and QC SPE extracts. The DI-HRMS spectra of the blank and upstream SPE extracts are dissimilar in composition and differ from the sewage outfall, downstream and QC extracts DI-HRMS spectra. In contrast, the sewage outfall, downstream and pooled QC all show the same equally spaced series of ions, indicative of an homologous series of compounds present in all three extracts. Comparing the upstream, sewage outfall and downstream DI-HRMS spectra, there is a clear compositional change when comparing the upstream and downstream riverine extracts which indicates that the sewage outfall has affected the DOM composition downstream of the point source.

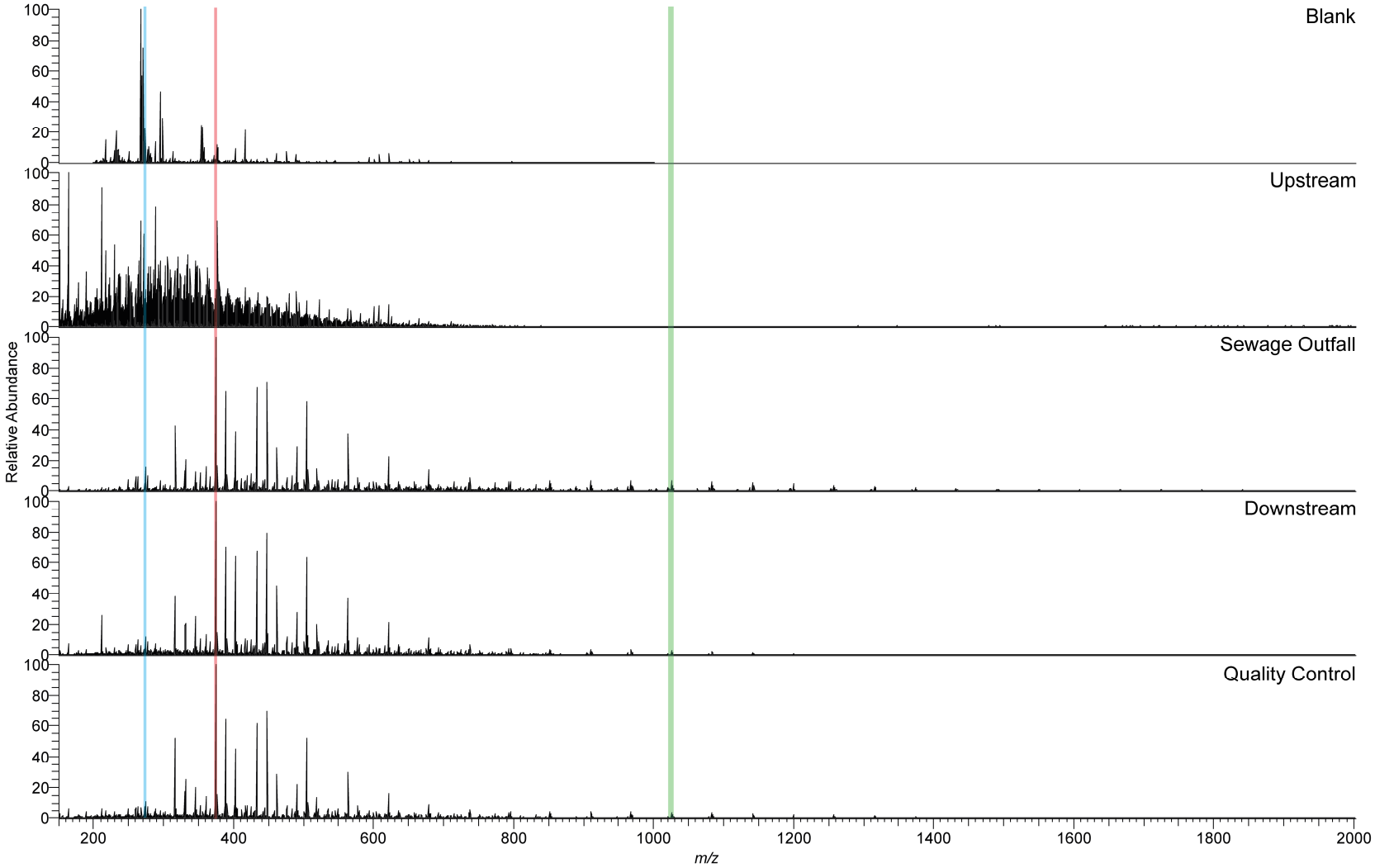


Figure 3.13. Representative DI-HRMS spectra of the blank, upstream, sewage outfall, downstream SPE extracts and pooled QC for the range *m/z* 150 to 2000. The blue, red and green boxes highlight the mass ranges, *m/z* 271.5 to 274.5, *m/z* 373.0 to 375.5, and *m/z* 1020 to1030 which are used in Figure 3.14.

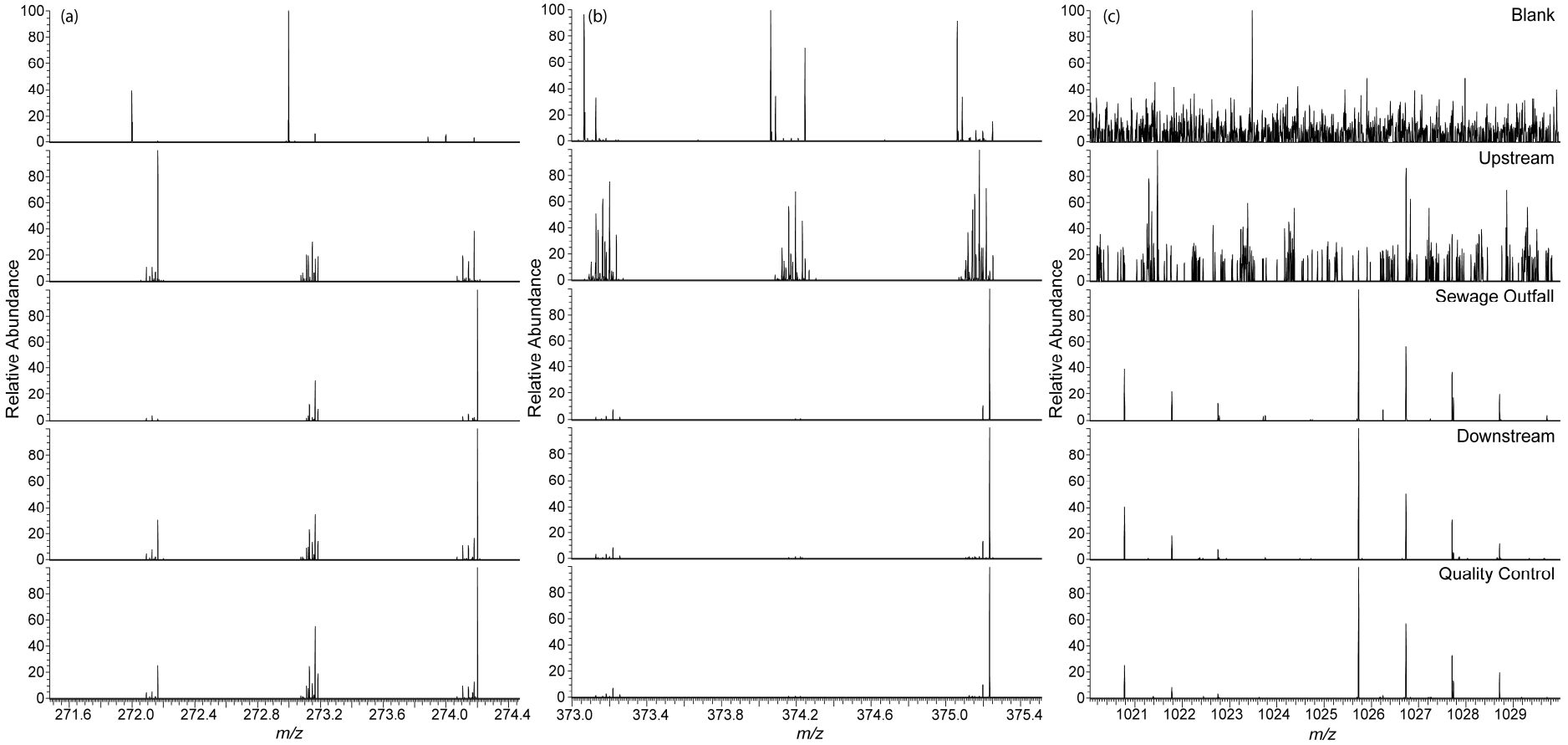


Figure 3.14. Narrow mass range DI-HRMS spectra for the blank, upstream, sewage outfall and downstream DOM SPE extracts and pooled QC, for the mass ranges: (a) *m/z* 271.5 to 274.5, (b) *m/z* 373.0 to 375.5, and (c) *m/z* 1020 to1030.

Figure 3.14 shows three narrower mass ranges within the DI-HRMS spectra which highlight some of the key differences and similarities between the composition of upstream, downstream and sewage outfall DOM SPE extracts. On reviewing the DI-HRMS spectra shown in Figure 3.13 in the range *m/z* 150 to 300 it is difficult to determine whether there are compositional differences between upstream, downstream and sewage outfall DOM SPE extracts, as the ions in this region of the spectra have a lower abundance in comparison to other ions in the DI-HRMS spectra. Figure 3.14(a) compares the ions in the mass range *m/z* 271.5 to 274.5, which shows that the ions in the upstream are present in the QC, downstream and sewage outfall spectra. However, there is an additional ion seen at *m/z* 274.201 which is not present in the upstream spectra, but is present in the sewage outfall, downstream and QC. This shows that there are compositional differences between the DOM SPE extracts which are not apparent when comparing large mass ranges of the DI-HRMS spectra. Therefore, a comprehensive comparison is not possible without comparing individual ions using narrower mass ranges within the DI-HRMS spectra.

The sewage outfall, downstream and QC DI-HRMS spectra of the DOM SPE extracts presented in Figure 3.13, shows that there is clearly a series of ions with a regular mass difference which is likely to be indicative of an homologous series of compounds. Figure 3.14(b) shows the range *m/z* 373.0 to 375.5 which contains the ion *m/z* 375.234. This ion has the highest abundance in the sewage outfall, downstream and QC DI-HRMS spectra and is found in one of the homologous series of molecular ions. Figure 3.14(b) shows that there is an additional ion at the same nominal mass *m/z* 375.198 in the sewage outfall, downstream and QC DI-HRMS spectra. Both the ions *m/z* 375.198 and *m/z* 375.234 are found in the upstream DI-HRMS spectra however, the abundance of these ions is much lower in the upstream DI-HRMS spectra. When comparing the DI-HRMS spectra shown in Figure 3.14(b) it is not possible to determine if the molecular ions seen in the upstream DI-HRMS spectra are found in the sewage outfall, downstream or QC. However, upon magnifying the DI-HRMS spectra of the sewage outfall, downstream and QC, the ions which are in the upstream are clearly present.

The DI-HRMS spectra in Figure 3.13 show that the sewage outfall, downstream and QC contain multiple ions between *m/z* 800 to 1600 compared to no ions seen in the upstream or blank DI-HRMS spectra. Figure 3.14(c) shows a 10 Da mass range *m/z* 1020 to1030 where there are ions in the downstream, sewage outfall and QC which are not found in the upstream or blank DI-HRMS spectra of the DOM SPE extracts. Based on the *m/z* and abundance of the

ions shown in Figure 3.14(c) the sewage outfall, downstream and QC DI-HRMS spectra contain two isotopic distributions, where the ions *m/z* 1020.774 and *m/z* 1025.729 are the monoisotopic ions for the two isotopic distributions. The ion *m/z* 1020.774 has two subsequent isotope ions *m/z* 1021.777 and *m/z* 1022.754, and the ion *m/z* 1025.729 has three isotope ions, *m/z* 1026.732, *m/z* 1027.709 and *m/z* 1028.713. This shows that the exclusion of elemental isotopes in molecular formula calculations which are commonly used to characterise the composition of ions in DI-HRMS spectra of DOM extracts and can be used to determine if an ion should be targeted for identification, could potentially misassign the molecular formula and effect further analyses (Koch *et al.*, 2007; Reemtsma, 2009; Gonsior *et al.*, 2011; Herzsprung *et al.*, 2014; Ruff *et al.*, 2015). Figure 3.14(c) shows that the DI-HRMS spectra of the upstream and blank have many ions with a random distribution, many unresolved and non-symmetrical profile peak indicating that these are noise. Therefore, this confirms as shown by the DI-HRMS spectra in Figure 3.13 that there are ions in the sewage outfall, downstream and QC which have a higher mass and are not in the upstream or blank DOM SPE extracts.

3.4.2.3 Summary of direct comparison of DI-HRMS spectra

The DI-HRMS spectra reveal remarkable complexity in the composition of the DOM extracts. For all the SPE extracts the spectra show multiple isobaric ions at each nominal mass. Overall, more than 1000 individual ions are present, the majority of which will correspond to individual molecular species. Results presented above illustrate the high degree of reproducibility of the approach as evidenced by the similarities between replicate analyses. The use of the mixed QC solution further confirms the stability of the Orbitrap. It is only when analysing a small mass range that individual ions can be compared. When comparing the DI-HRMS spectra using these smaller mass ranges, it is difficult to compare all water SPE extracts and replicates together. The figures above displaying the smaller mass ranges of the DI-HRMS spectra, as shown by the highlighted regions only represent a small fraction of each the DI-HRMS spectra of the SPE water extracts. A comprehensive direct comparison aimed at determining all compositional differences in DOM of the water samples studied herein is simply unfeasible; it would require the alignment of 25 DI-HRMS spectra followed by systematic analysis of 3-10 Da mass ranges for the full recorded range *m/z* 150 to 2000. Furthermore, once an ion(s) has been found which changes in abundance or presence/absence in extracts, it is impossible to determine the significance of such changes. Multivariate statistical analysis, such as PCA,

offers a practical way of exploring differences in DI-HRMS data of extracts from different sources.

3.4.2.4 Pre-processing of DI-HRMS spectra prior to statistical analysis: ion picking and alignment

Prior to PCA, ion *m/z* values and intensities need to determined from the DI-HRMS spectra of all of the DOM SPE extracts of water. Different ion picking (commonly referred to as "peak picking") and alignment parameters have been used in previous studies to align DI-HRMS spectra of DOM extracts, including alignment by molecular formula (Chen *et al.*, 2014) and exact mass (Li and Minor, 2015). Using an approach of aligning by molecular formula would require the comprehensive assignment of molecular formula prior to analysis and would have the same disadvantages as discussed in Chapter 1. The method of Li and Minor (2015) involved an in house written script which is not publicly available. The R package XCMS was designed for peak picking and alignment of high resolution mass spectra either with prior chromatographic separation or directly infused (Smith *et al.*, 2006; Tautenhahn *et al.*, 2008; Benton *et al.*, 2010). The alignment of ions using XCMS across the different DI-HRMS spectra uses the accurate mass of ions. Therefore, determining the variation in mass accuracy is important to correctly align ions.

An ESI positive ion calibration standard (Pierce™ LTQ ESI Positive Ion Calibration Solution) was run every 5 analyses in a sequence. The scan-to-scan variability of the mass accuracy of each ion in the standard was assessed at the start and end of the analyses, using the mass check function in the Orbitrap software and showed a scan-scan fluctuation range of 3.2 ppm. Figure 3.15 shows the changes in mass accuracy of the standard mixture across the analyses of the DOM SPE extracts. There was a drift of 1.8 ppm in the mass accuracy over the analytical run. Therefore, alignment of ions across all the analyses of the DOM SPE extracts DI-HRMS spectra was set at 5 ppm. The ion picking was carried out as described in Section 2.3.2.1 3237 individual ions were picked above a signal-to-noise ratio of 5 and aligned across all extracts.

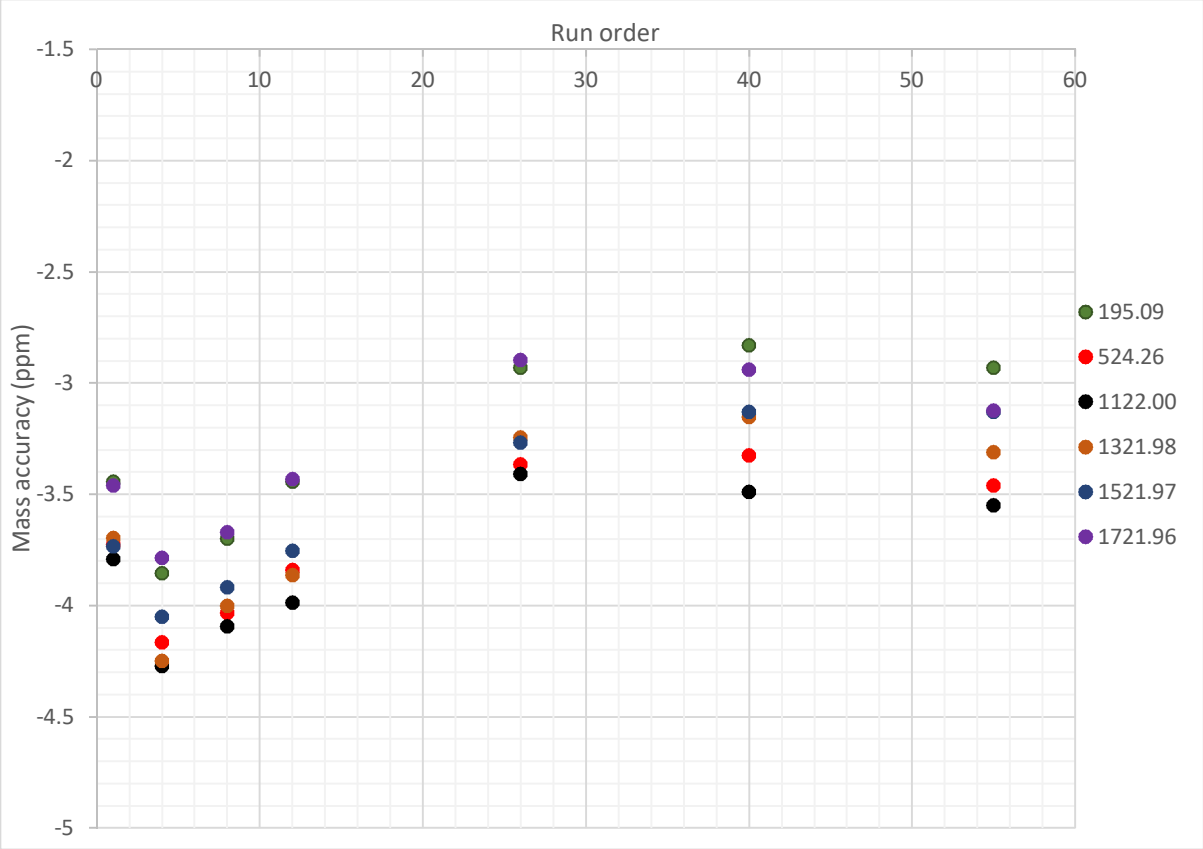


Figure 3.15. The changes in mass accuracy of Pierce™ LTQ ESI Positive Ion Calibration Solution standard ions across run show that the mass accuracy of the ions across the run varied by 1.8 ppm.

3.4.3 Statistical analysis of the DI-HRMS spectra of the SPE extracts

The statistical analyses were carried out using Mass Profiler Professional (Agilent, USA) for both multivariate and univariate statistics. The intensity values were first normalised using $\log_2$. Three statistical tests were then applied: (i) PCA, (ii) hierarchical cluster analysis combined with a heatmap, (ii) Kruskal-Wallis one-way analysis. PCA was used to assess if there was a significant difference between the composition of the upstream, downstream and sewage outfall SPE extracts.

The PCA in Figure 3.16 highlights the differences in composition of the DI-HRMS spectra of the QC, upstream, downstream, sewage outfall, and blank DOM SPE extracts comparing the principal components (PC) 1 and 2. The QC was used to assess whether there was any significant differences caused by analytical performance, ion picking, alignment or normalisation. The QC is clustered in line with the upstream, downstream and sewage outfall, which was expected as it is a mixture of all three SPE extracts. The QC clusters between the downstream and the sewage outfall because as shown when comparing DI-HRMS spectra

(Figure 3.13), the QC contains molecular ions from the sewage outfall which are not present in the upstream extracts. The clustering of the QC replicate analyses shows there is some variance in the analysis, however, it is not as large as the difference between the upstream, downstream and sewage outfall composition. This demonstrates that there are no significant differences caused by the analytical variance or pre-processing, which might affect the statistical analysis.

The PCA in Figure 3.16 shows all replicates cluster in their respective groups confirming that the minor differences observed between the analyses of replicate SPE extracts in the direct comparison of the DI-HRMS spectra discussed in Section 3.4.2.1 are not as significant as the overall differences in composition of DOM SPE extracts collected from different sites. Thus, the differences in the PCA are not due to analytical or extraction variance. The vector PC1 explains the greatest fraction of the total variance 57.52 %, which differentiates the composition of the blank HPLC water extract and all the DOM SPE extracts, confirming the manual assessments of the DI-HRMS spectra discussed in Section 3.4.2.1. PC2 explains the next highest fraction of total variance 12.19 %, which as expected shows separation of the DI-HRMS spectra of the upstream, downstream, QC and sewage outfall, wherein the latter is least similar to the composition of the upstream SPE. As expected, the pooled QC and downstream SPE DOM extracts plot between the upstream and sewage works DOM, since both are a mixture of background riverine and the sewage works DOM.

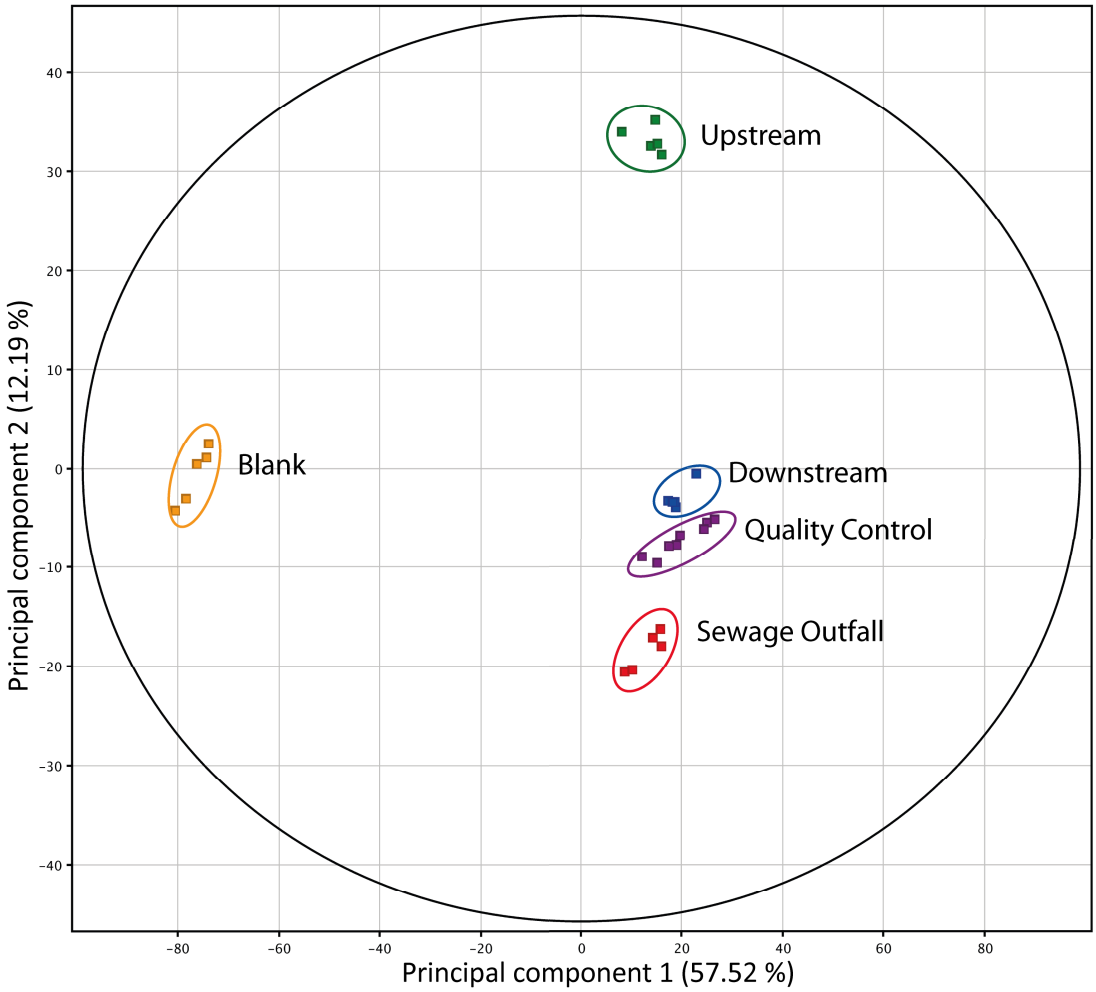


Figure 3.16. PC1 vs. PC2 scores plot of the DI-HRMS spectra of the upstream DOM (■), downstream DOM (■), sewage outfall DOM (■), QC (■) and blank (■) SPE extracts.

The PCA in Figure 3.17 compares the DI-HRMS spectra upstream, downstream and sewage outfall DOM SPE extracts. This was assessed to determine whether the inclusion of the blanks and QC were influencing the separation observed when comparing the upstream, downstream and sewage outfall DI-HRMS spectra of the DOM extracted using SPE. PC1 explains the highest fraction of the total variance 46.9 %, which shows the DOM compositions DI-HRMS spectra of the sewage outfall and upstream DOM SPE extracts are least similar. The composition of downstream extract DOM plots between the sewage outfall and upstream SPE extracts as expected, because the downstream DOM should be a mixture of the riverine and sewage outfall's DOM composition. The PC2 shows that there is variance between the replicate samples however this explains only 11.35 % of the total variance which is four times smaller than the fraction of the total variance explained by PC1. This agrees with what was observed when directly comparing the DI-HRMS spectra which showed that there were clear compositional differences between sewage outfall and upstream DI-HRMS spectra, and the

downstream clearly contained ions in both the upstream and sewage outfall DI-HRMS spectra. Furthermore, when comparing the DI-HRMS spectra of the replicate SPE extracts smaller changes between the DI-HRMS spectra were highlighted, however the PCA shows that these were smaller than the differences between the composition of the DOM extracted from the 3 water sources.

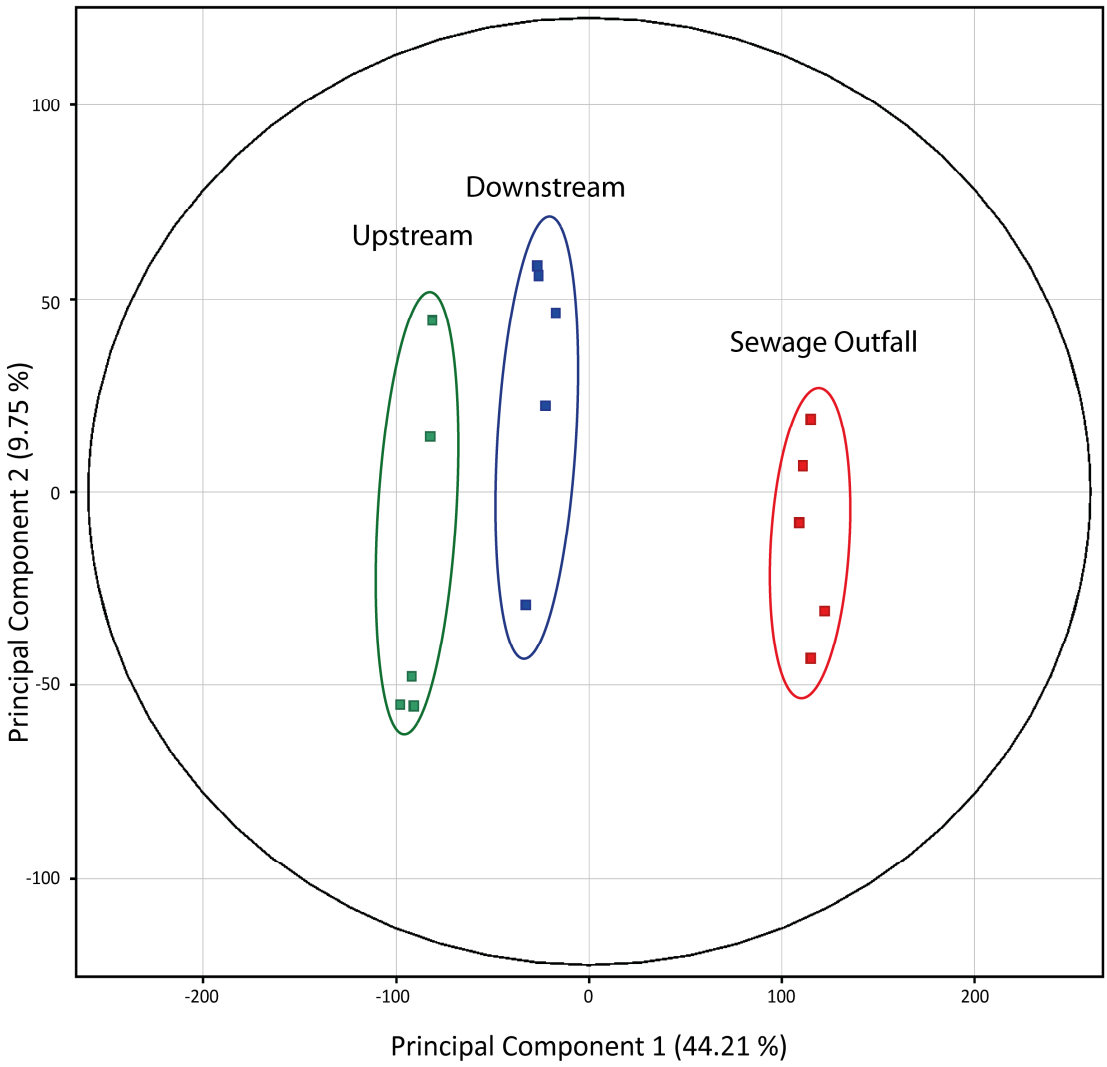


Figure 3.17. PC1 vs. PC2 scores plot of the DI-HRMS spectra of the upstream (■), downstream (■), sewage outfall (■) DOM SPE extracts.

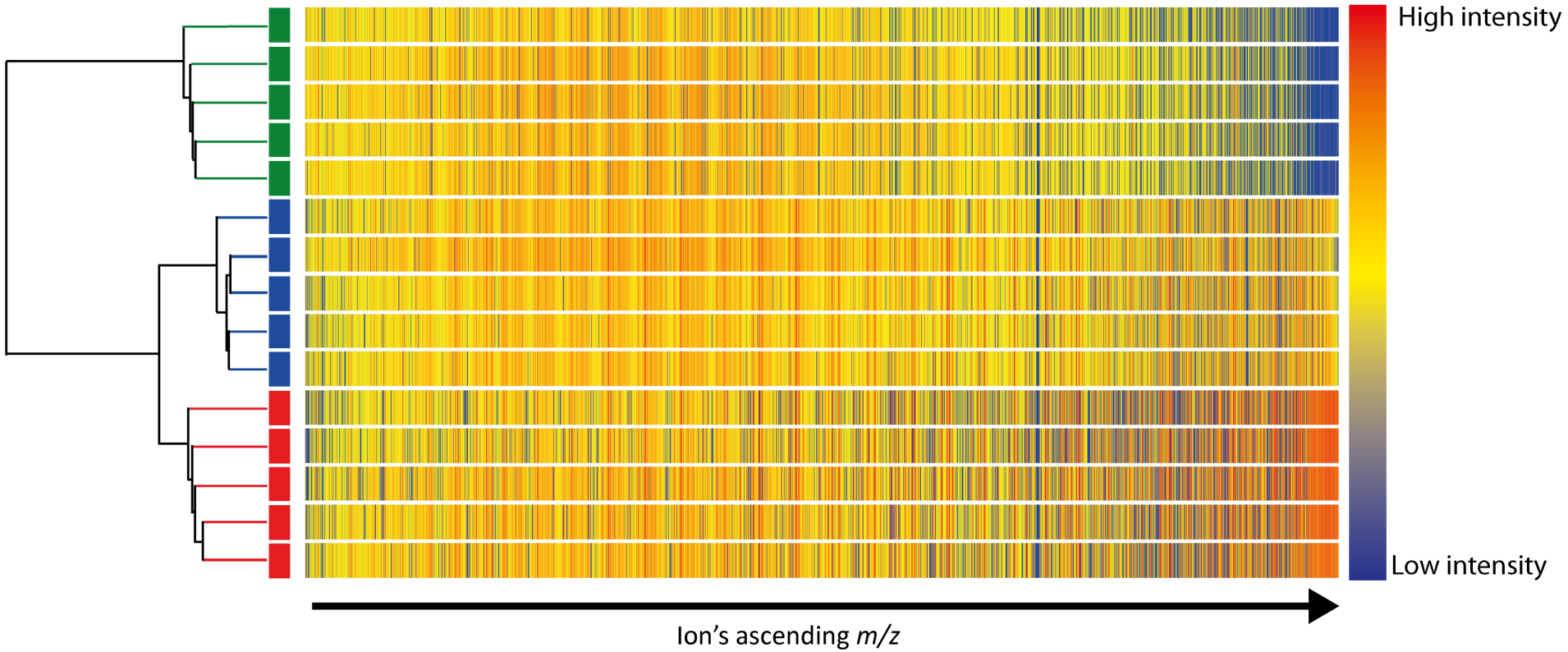


Figure 3.18. Hierarchical cluster analysis (left) of DI-HRMS spectra of the upstream (■), downstream (■), sewage outfall (■) SPE extracts. Heatmap (right) shows the ions in each of the DI-HRMS spectra arranged by mass on the x axis. Comparison of the $\log_2$ of the intensity of the ions represented by colour with higher intensity hotter (red) and lower intensity colder (blue).

Figure 3.18 shows the hierarchical cluster analysis of the DI-HRMS spectra of the upstream, sewage outfall and downstream SPE extracts combined with a heatmap of the ions in the DI-HRMS spectra. The heatmap represents all 3237 ions picked from the DI-HRMS spectra arranged in order rom lowest to highest mass (left to right) and compares the $\log_2$ normalised intensity of ions by colour. Higher intensity ions are represented in hotter colours (red) and lower intensity in colder colours (blue). The hierarchical cluster analysis shows that the sewage outfall and downstream are more similar in composition as these separate lower on the dendrogram than the upstream. Furthermore, all replicates cluster in their three respective sample groups showing consistency between their extract compositions.

Manual assessment of the DI-HRMS spectra, directly based on individual ions, required the investigation of narrower mass ranges, making comprehensive assessment of such complex spectra in this way impossible. The heatmap provides an effective way of visualising changes in ion abundances between DI-HRMS spectra of SPE extracts and replicates as each ion is represented by a line across the spectra. The differences in colour provides an effective way of identifying ions which vary in abundance between replicates and sample groups, making it easier to identify which masses are different or similar in composition between replicates and sample extracts.

Comparing the heatmap of the upstream and sewage outfall DI-HRMS spectra reveals many ions which vary in intensity between the two locations making it difficult to specify ions which differentiate between the DI-HRMS spectra. However, ions which have higher intensity in the all extraction replicates of the DI-HRMS spectra of the sewage outfall will correspond to compounds originating from this source. The higher mass range shows a clear cluster of ions which are have a lower intensity in the upstream DI-HRMS spectra, high intensity in the sewage outfall DI-HRMS spectra and more neutrally represented in the downstream DI-HRMS spectra. This reflects what was seen in the DI-HRMS spectra shown in Figure 3.14, which compared ions in the range *m/z* 1020 to 1030. Furthermore, this exemplifies the behaviour anticipated for compounds originating from the sewage outfall, i.e. compounds from sewage outfall will be diluted by the river flow and therefore should be reduced in abundance in the downstream DI-HRMS spectra.

To determine which ions are most significant in differentiating between DI-HRMS spectra of the point source and background river DOM univariate statistics were used to compare the upstream and sewage outfall molecular fingerprints as these were found to be least similar in

composition using PCA. First the data was assessed to check to determine if the abundance values were normally distributed for each ion. The kurtosis and skewedness of the distribution the abundances of each of the ions were calculated to assess pointedness and skew. This was to determine whether the data was normally distributed. However, the variance of both these metrics meant the data was not consistently normally distributed therefore a non-parametric test was deemed the most suitable to further compare the data. Therefore, Kruskal-Wallis analysis was used to compare all the ions in the upstream and sewage outfall DI-HRMS spectra using Mass Profiler Professional (Agilent, USA). Kruskal-Wallis analysis is a non-parametric test which can be used to compare by rank 2 or more conditions (ions) and their measured values (intensities) to determine if these are significant in differentiating between the sample groups (extracts). Ions with a p value < 0.005 were determined to be statistically significant in differentiating between the extracts. Table 3.2 shows the number of ions found at common p value significance thresholds, furthermore whether the ion's abundance was increasing or decreasing in the sewage outfall DI-HRMS spectra when compared to the upstream DI-HRMS spectra abundance values. The number of ions predicted by chance for each p values significance threshold was also calculated to ensure that the results were not a result of chance.

Table 3.2. p values of ions from the Kruskal Wallis analysis of the sewage outfall and upstream samples over a range of commonly used p value thresholds of statistical significance and the number of ions expected by chance.

p value	p all	p<0.05	p<0.02	p<0.01	p<0.005
Number of ions	3237	2358	2087	1846	1451
Number of ions decreasing	2353	1735	1498	1284	941
Number of ions increasing	884	623	589	562	510
Expected by chance	N/A	161	64	32	16

Table 3.2 shows that for each of the significance thresholds there are more ions than predicted by chance. To assess which ions, originate from the sewage outfall, only ions which increased intensity in the sewage outfall DI-HRMS spectra were retained for further analysis. Furthermore, only ions which had a p value < 0.005 were prioritised for further analysis. Figure 3.19 highlights within the DI-HRMS spectrum of the sewage outfall DOM SPE extract the 510 ions which have a p value < 0.005 and increase in abundance in the sewage outfall. These ions clearly include the homologous series of ions observed in the sewage outfall mass spectra and were absent from the upstream spectra and includes the ions highlighted with a higher *m/z*

which were also absent from the upstream DI-HRMS spectra. However, there are other ions whose behaviour is less obvious when comparing the different mass spectra. In such cases the Kruskal-Wallis analysis provides an effective data reduction step to determine which ions to prioritise for further analysis. This improves upon the current empirical approaches used for untargeted analysis of micropollutants in the environment as it eliminates the need for manual analysis and specifically the need for molecular formula determination and making decisions as to which molecular specific to target as potential micropollutants.

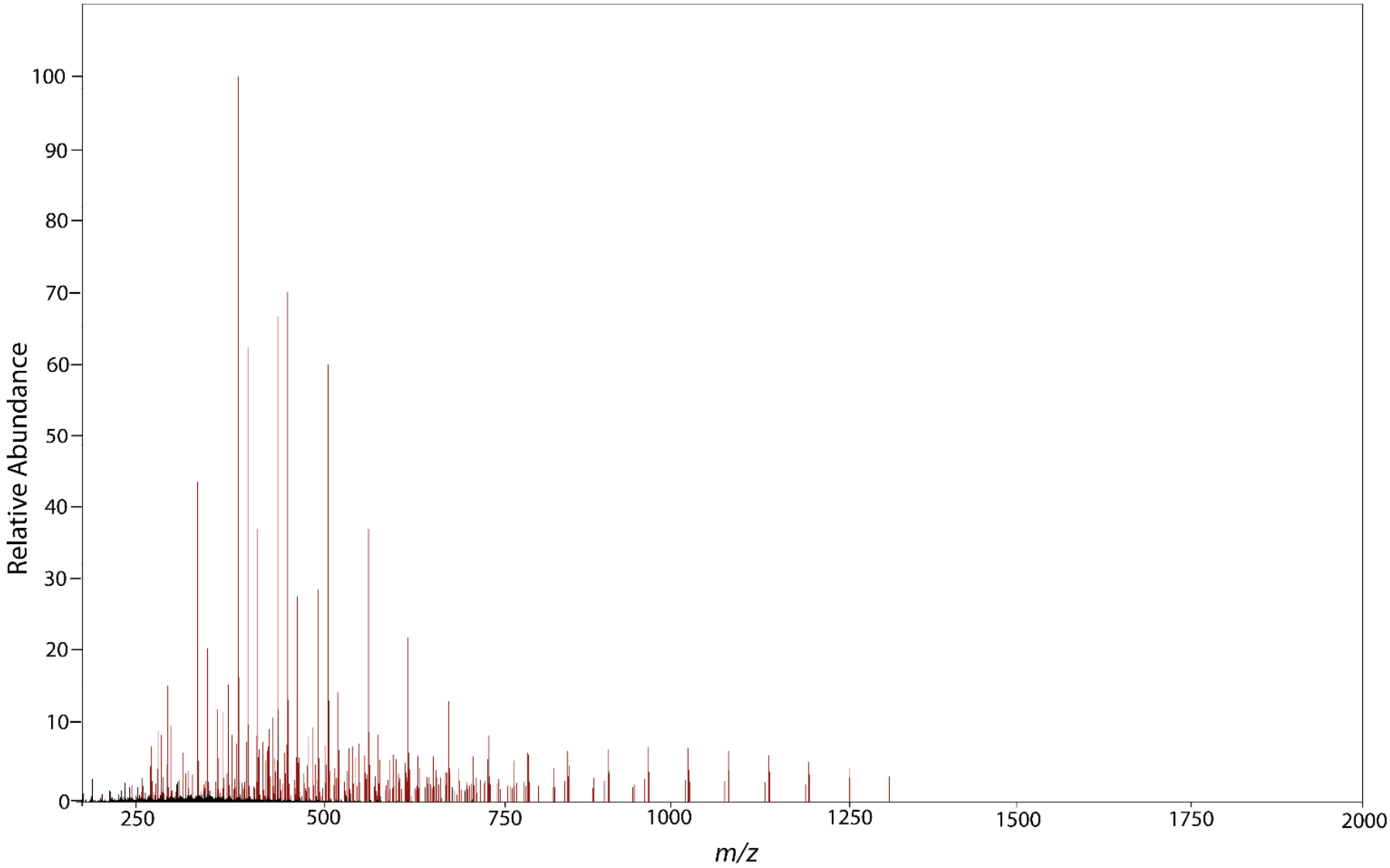


Figure 3.19. DI-HRMS spectra of the sewage outfall extract with the 510 ions with a p value < 0.005 and which increase abundance in the sewage outfall highlighted in red.

3.4.4 Tentative identification of homologous series of molecular adduct ions

As shown when comparing the DI-HRMS spectra in Figure 3.13 one of the key differences seen between the upstream DI-HRMS spectra and the downstream, QC and sewage outfall DI-HRMS spectra was the presence of multiple series of ions with a common mass defect. By comparing the mass difference of these it was found that the ions had a difference in mass of 58.041 Da which is known to be indicative of polypropylene glycol (PPG). To further investigate this a pattern matching algorithm was written in R as discussed in Section 2.3.2.2 to find sequential ions related by a specified mass difference.

Using the pattern matching algorithm, series with more than 5 sequential ions where the difference of the masses of the ions were within 10 ppm of 58.041 Da were selectively extracted from the sewage outfall DI-HRMS spectra. Eighteen sequences containing 177 individual ions were found to have 58 Da spacing. Of these 177 ions, 157 had a p value < 0.005 and increased ion abundance when comparing the sewage outfall and upstream DI-HRMS spectra. All the sequences had a normal distribution of ion abundance, four of which are highlighted in Figure 3.21. The 18 sequences are summarised in Table 3.3.

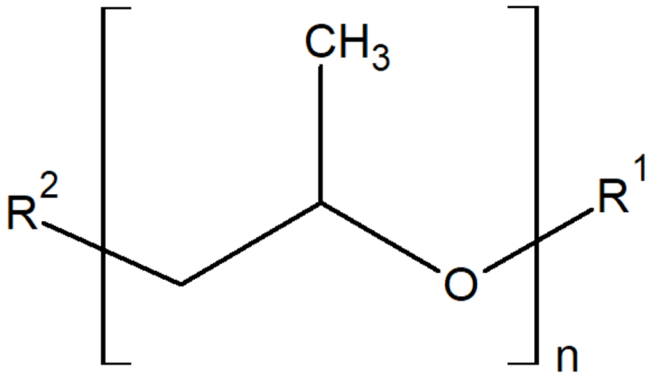


Figure 25. Generic PPG structure with the end groups represented as R^1 and R^2

PPG polymers have a repeating monomer unit of C_3H_6O, however, end groups of these polymers can vary. The synthesis of PPG is initiated using a compound with an active proton group and depending on the initiator used the end group of the polymer will vary. This reaction is quenched with acid and by changing the initiator the properties, rate of reaction and molecular weight distributions of PPG can be controlled (Steiner *et al.*, 1964; Huang *et al.*, 2002). PPG has several uses because of its high solubility in water and low toxicity, these include: food flavourings, surfactants, antifreeze, paints and coatings (Steiner *et al.*, 1964;

Huang *et al*., 2002). PPG has been detected in environmental surface waters using C_{18} SPE extraction, derivatisation of the end groups of PPG with the fluorophore 1-napthyl iso-cyanate followed by reverse phase HPLC with fluorescence detection (Rychłowska *et al.*, 2003).

Assuming all ions in the 18 series are PPG, the difference in mass between the series could be due to a difference: of isotopic composition, end group of the PPG molecule or a difference in the adducting species during HESI. Therefore, by subtracting the mass of repeating monomer units propylene glycol ($-C_3H_6O-$) from the mass of the ion the difference can be used to calculate the formulae of the adduct, isotope and end groups. Equation 3.2 shows the calculation for the mass difference (M_d) which relates to the mass of the adducts, isotopes and end groups, where M_x is measured exact mass (*m/z*) recorded in the DI-HRMS spectra, M_{pg} is the mass of the monomer of PPG (58.04186 Da).

$$\left(\frac{M_x}{M_{pg}} - \left\lfloor\frac{M_x}{M_{pg}}\right\rfloor\right) \times M_{pg} = M_d \qquad (3.3)$$

Without further analysis of the structures of these ions it is not possible to determine which elements of the molecular formula are due to the adduct or end group and whether these series are PPG. However, using the formula of the isotopes, it is possible to determine which are the monoisotopic PPG series and the corresponding ^{13}C PPG series. Table 3.3 shows it is possible to determine for the monoisotopic PPG series S1, S2 and S3 the corresponding ^{13}C PPG series are S6, S8 and S16, respectively.

Table 3.3. PPG sequences with their highest and lowest masses in the series, the range of the number theoretical monomer units, the range of p values for each series, molecular formulae of end groups, adducts and possible isotopes.

	Mass range (*m/z*)	**Monomer units**	**Max p value**	**Min p value**	**End group and adduct†**	**Isotope**
PPG S 1	445.3128- 1315.9393	7 - 22	0.0032	0.0013	$NaCH_4$	
PPG S 2	273.1669 -1027.7093	4 - 17	0.0027	0.0013	NaH_2O	
PPG S 3	251.1846 - 947.6849	4 - 16	0.0013	0.0013	H_3O	
PPG S 4	215.0851 - 737.4638	2 - 11	0.0035	0.0013	$NaC_2H_4O_3$	
PPG S 5	287.1461 - 809.5214	4 - 13	0.0020	0.0013	NaO_2	
PPG S 6	736.5244 - 1258.9007	12 - 21	0.0035	0.0013	$NaCH_4$	^{13}C
PPG S 7	289.1257 - 753.4587	3 - 11	0.0016	0.0013	$NaC_2H_4O_4$	
PPG S 8	332.2117 - 796.5455	5 - 13	0.0013	0.0013	NaH_2O	^{13}C
PPG S 9	730.5659 - 1194.8996	12 - 20	0.0041	0.0013	NCH_8	
PPG S 10	259.1511 - 665.4429	4 - 11	0.0013	0.0013	Na $C_2O_2H_4$	
PPG S 11	303.1413 - 709.4328	5 - 12	0.0013	0.0013	NaO_3	
PPG S 12	329.1928 - 735.4849	5 - 12	0.0013	0.0013	NaO	
PPG S 13	333.1514 -739.4431	4 - 11	0.0016	0.0013	$NaCH_2O_4$	
PPG S 14	371.2395 - 777.5315	6 - 13	0.0013	0.0013	Na	
PPG S 15	384.2947 - 790.5865	6 - 13	0.0013	0.0013	NH_6O	
PPG S 16	426.3134 - 832.6056	7 - 14	0.0013	0.0013	H_3O	^{13}C
PPG S 17	317.1564 - 665.4069	5 - 11	0.0016	0.0013	$NaCO_3H_2$	
PPG S 18	359.2034 - 707.4534	5 - 11	0.0016	0.0013	$NaCO_2H_2$	

† Without further structural characterisation it is not possible to differentiate between the formula of the adduct or end group solely using the exact mass. Therefore, the collective molecular formula of the adduct and end group is used in Table 3.3.

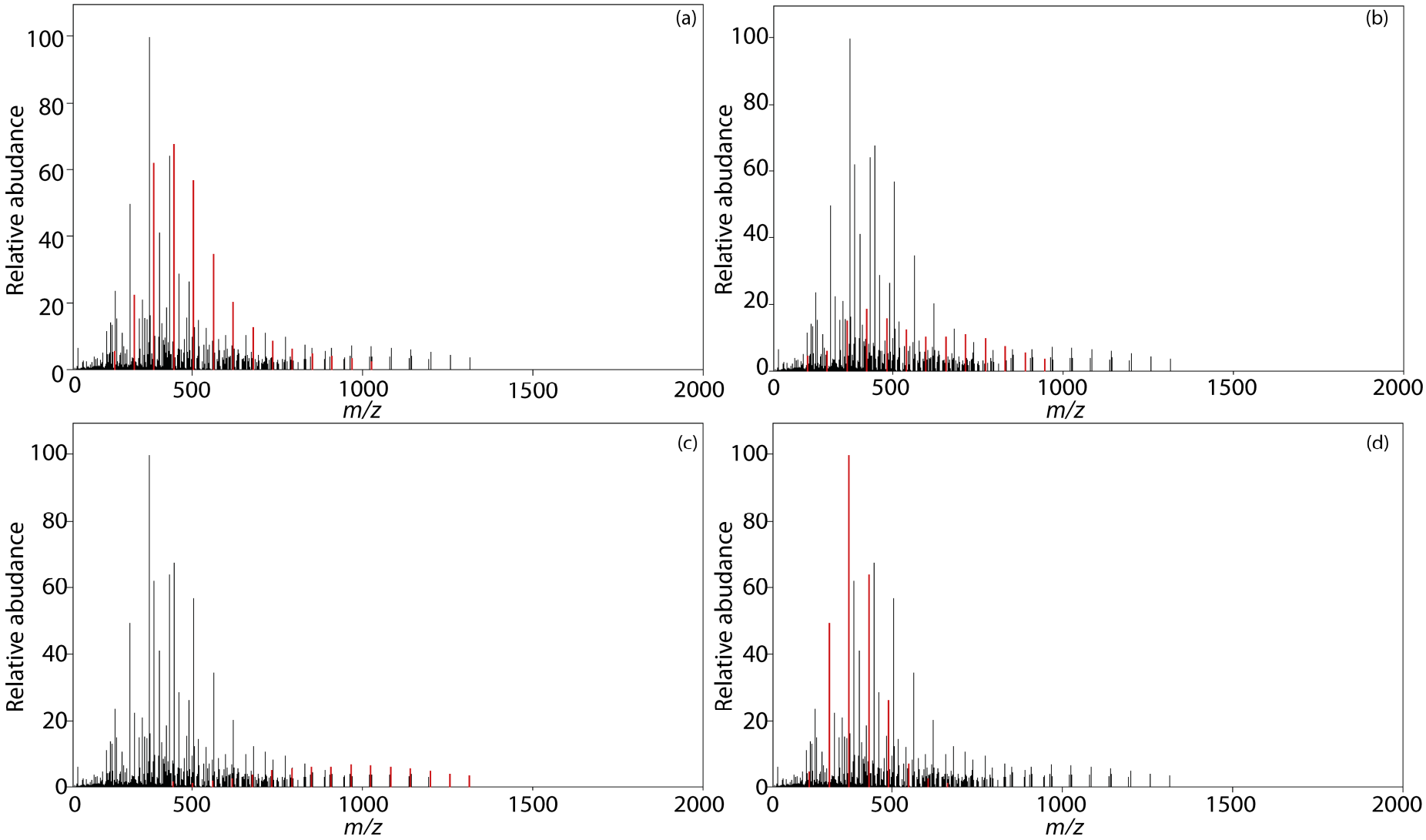


Figure 3.21. DI-HRMS spectra of DOM SPE extract from the sewage outfall water. Four of the PPG series, from table 3.3 are highlighted in red, showing (a) PPG series 2, (b) PPG series 3, (c) PPG series 1, and (d) PPG series 10.

3.5 Summary

The DI-HRMS spectra show how complex the SPE extracts from the different water sources are and the impact that sewage outfall has upon the riverine composition (Aim i). However, a limitation of the method presented is that changing the extraction method, analytical method or analytical parameters (i.e. the ionisation method or polarity) may result in the analysis of a different fraction of the DOM, thereby yielding different results or highlighting different discriminating components. This shows the benefit of initially using DI-HRMS to analyse and compare the extracts collected, as this is a quick analytical method (1 – 3 min) in comparison to HPLC-MS or GC-MS (>15 min). Thereby allowing statistical datasets to be collected quickly and use multivariate statistics to initially determine if there is a compositional difference between the fraction extracted before more time-consuming analytical methods are used to interrogate the composition of the extracts.By using multivariate statistical methods, it was possible to show the consistency between replicate extracts and the differences between extracts from different sampling locations (aim ii, iii & iv).

As discussed in the introduction the most laborious process is identification of individual compounds and as we have seen from the DI-HRMS spectra shown throughout this chapter there are thousands of individual ions which are potentially compounds contained within these extracts, making identification prohibitively time consuming. The Kruskal-Wallis analysis provides an essential data reductive step (aim v). Certain ions are highlighted which are found to be statistically significant. These are potentially originating from the sewage outfall and therefore are likely to be anthropogenically derived micropollutants. However, there is the possibility that other ecotoxicological compounds; not extracted, nor detected using this analytical method, or found to be statistically significant using this method maybe missed. This method provides an unbiased starting point in determining the composition of DOM from point sources.

By using the accurate mass of the statistically significant ions it is possible to determine the molecular composition of these compounds and postulate an identity. However, solely using their accurate mass and molecular formulae would not unequivocally identify these compounds. As shown, the complexity of these DI-HRMS spectra have multiple ions resolved ions within a 1 Da range. Even though the Orbitrap can have a low isolation width (<1 Da) multiple ions would be isolated for further MS^2 experiments. This would result in a chimeric product ion spectra making it difficult to identify individual ions. Therefore, chromatographic

separation is essential to identifying these compounds and will be discussed further in Chapter 4.

3.6 Conclusions

This chapter presents the results from the use of ESI-Orbitrap HRMS using DI-HRMS in the analysis of SPE extracts of DOM in river water. The major findings are that:

(i) The DI-HRMS molecular 'fingerprints' of the DOM extracts of river water obtained using SPE reveal the exceptional complexity of their composition. (aim i)

(ii) The DI-HRMS spectra of the DOM extracted from the upstream, downstream and sewage outfall water, illustrates how a point source can dramatically alter the composition of the riverine DOM. (aim iii)

(iii) Manual assessments of the DOM composition, while revealing specific spectral features driving differences in DOM composition, emphasise the need to use chemometric statistical methods to process data sets of this complexity.

(iv) PCA analysis of the DI-HRMS spectra was readily able to resolve the different DOM sources, including the in-stream mixing. (aim ii & iv)

(v) The use of a pooled QC confirmed that the separation between the sample groups was not due to analytical variance, biases in ion picking and alignment and/or the normalisation of the ion intensities.

(vi) Hierarchical cluster analysis showed that all sample replicates cluster in their respective sample groups. Furthermore, it also showed that the composition of the downstream DI-HRMS spectra was more similar to the sewage outfall than the upstream SPE extracts, confirming the importance of the point source to the overall DOM. (aim ii & iv)

(vii) Heatmapping facilitated visualisation of the changes in the abundances of ions between DI-HRMS spectra. (aim iii)

(viii) Comparison of the sewage outfall and upstream DI-HRMS spectra using Kruskal Wallis analysis provided a critical statistical data reduction step to identify the most important molecular species driving the differences in composition between the DOM extracts. Unlike previous untargeted analysis studies this does not require the prior determination of the molecular formula of individual ions and the determination of which molecular formulae make an ion significant (Gonsior *et al.*, 2011; Ruff *et al.*, 2015; Hollender *et al.*, 2017). (aim v)

(ix) At least eighteen homologous series were identified with normal distributions of ions separated by 58 Da, tentatively identified as PPG in the sewage works outfall DOM.

Overall, the application of HESI-Orbitrap HRMS to riverine DOM extract has demonstrated a high level of chemical complexity which defies manual assessment and interpretation. This has been partially resolved using multivariate and univariate statistical analyses to assess which molecular species vary between sources. This provides an important first step in the comprehensive assessment of micropollutants in the environment, allowing prioritisation of molecular species contributed from a point source.

The following chapter will demonstrate the next step, namely the interrogation of the molecular species using HPLC-HESI-Orbitrap HRMS.

Chapter 4

Identification of compounds from Chew Stoke sewage works contributing to riverine DOM using HPLC-Orbitrap-HRMS

Chapter 4. Identification of compounds from a sewage works contributing to riverine DOM using HPLC-Orbitrap MS

4.1 Introduction

As discussed in Chapter 1 the identification of micropollutants is critical to understanding the impact that pollution is having on the environment. Without the identification of these micropollutants, it is not possible to determine their ecotoxicity, nor monitor their concentration in the environment. Previous targeted studies identify compounds within sewage effluent or DOM using a predetermined list of compounds because they have either been previously identified or predicted to be present within a water sample. (Andresen *et al.*, 2004; Gros *et al.*, 2009; Loos *et al.*, 2012; Verlicchi *et al.*, 2012). These studies use optimised extraction methods designed to isolate and concentrate the target analytes from other compounds within the water samples (e.g. Boyd-Boland *et al.*, 1996; Andresen *et al.*, 2004; Kosonen and Kronberg, 2009; Verlicchi *et al.*, 2012; Santos *et al.*, 2013). GC or HPLC are used to chromatographically separate compounds prior to MS and therefore use both the product ion spectrum and retention time to identify the compounds of interest. The peaks in the TIC or EIC can be used alongside known reference standards to quantify the target analytes. These methods can be used to determine the load of a compound in a source or waterbody (e.g. Sabik *et al.*, 2000; Kosonen and Kronberg, 2009; Loos *et al.*, 2012; Santos *et al.*, 2013; Golovko *et al.*, 2014; Ruff *et al.*, 2015). However, these studies are predicated on being able to know which compounds to identify prior to analysis. As targeted methodologies only identify analytes of interest, other compounds within an extract or water sample are not analysed. Therefore, these targeted methods do not comprehensively assess the DOM in a system. Additionally, a single compound may break down into several transformation products, which may still be biologically active. Therefore, only analysing the target compound will underestimate the effect that a compound is having on the aquatic system (Haddad and Kümmerer, 2014).

Although previous studies have attempted to identify compounds alongside targeted studies prioritising ions based on peak size or molecular formula (Gonsior *et al.*, 2011; Ruff *et al.*, 2015). However, the limitation of these approaches is that the size of a peak is not necessarily an indication of it's significant in an extract, and not all micropollutants have a molecular composition which will differentiate them from the other compounds in DOM.

Previous untargeted studies have used statistics to determine if there is a difference between DI-HRMS spectra of DOM extracts from different environmental settings or extraction methods (Dubinenkov *et al.*, 2015; Li and Minor, 2015). However, these studies do not aim to identify the particular molecular ions which differentiates between extracts beyond a molecular formula characterisation (e.g. Chen *et al.*, 2014; Dubinenkov *et al.*, 2015; Li and Minor, 2015; Li *et al.*, 2017). Chapter 3 showed that it was possible to use a statistical approach to analyse point sources in comparison to the receiving riverine DOM. Kruskal-Wallis analysis was used to determine the significance of an ion in discriminating between these extracts. As a result, the most significant ions which were found to increase in abundance in the sewage outfall were selected as a target mass list to identify.

Many micropollutants, including pharmaceuticals, illicit drugs, and surfactants have been shown to originate from sewage effluent (e.g. Gros *et al*., 2009; Loos *et al.*, 2012; Santos *et al.*, 2013; Ruff *et al.*, 2015). An EU wide monitoring survey conducted by Loos *et al*. (2012) looked at sewage effluent targeting 161 priority pollutants and found 37 compounds in more than 85 % of sewage treatment works studied. Some of the compounds identified in previous studies of sewage outfalls would be expected to be statistically significant because they are also coming from Chew Stoke sewage outfall. Finding some of these micropollutants would show that this method highlights a relevant list of compounds. Furthermore, there may be additional compounds identified which have not been previously found in sewage effluent.

As shown from the DI-HRMS spectra in Chapter 3 DOM is an extremely complex mixture of molecular ions. Multiple clusters of ions were observed at 1 Da intervals in all DI-HRMS spectra. Despite advances in HRMS, one of the limitations is the isolation width in which a single molecular ion can be isolated for subsequent fragmentation experiments (Viant and Sommer, 2013). For example, using Orbitrap mass spectrometers, ions can be isolated within a predetermined mass window which can be set as low as a 0.4 Da. However, the smaller the isolation width the greater the loss in sensitivity. Because of the complexity of DI-HRMS spectra of DOM extracts an isolation width of 0.4 Da would result in the isolation of additional isobaric ions and structural isomers. A study by Gonsior *et al.*, (2011) isolated and fragmented ions from the DI-HRMS spectra. However, in comparison to the DI-HRMS spectra of the Chew Stoke water extracts, Gonsior *et al*. (2011) spectra were much less complex allowing for isolation of ions directly from the DI-HRMS spectra.

Therefore, to identify compounds chromatographic separation is required prior to MS/MS fragmentation. HPLC-MS analysis has been widely used in target compound analysis for the quantification and identification of compounds in DOM. A reverse phase C_{18} column was selected for this research because such columns are commonly used in previous studies of sewage treatment works to identify compounds (e.g. Kasprzyk-Hordern *et al.*, 2007; Sheng *et al.*, 2014; Ahrens *et al.*, 2015; Ruff *et al.*, 2015).

By using HPLC-MS instead of DI-HRMS the SPE extracts can be compared in more detail. Firstly, this analysis will allow the comparison of the concentration of compounds in each extract. Secondly, additional detection methods could be used prior to analysis using MS providing information about the structure of analytes. Finally, an ion in DI-HRMS spectra may be the result of multiple structural isomers, which may be separated by HPLC-MS.

4.2 Aims

The overall aim of this study is to assign specific structures to ions found to differ between the DI-HRMS spectra of the river background and sewage works DOM extracts investigated in Chapter 3, specific aims include:

(i) Further analysis DOM SPE extracts using HPLC-MS.

(ii) Comparison of the differences in peak areas of individual ions in the HPLC-MS analysis of different SPE extracts.

(iii) Determine the retention time of ions with the same exact mass for further analysis. Using the exact masses of the ions determined to be significant by the Kruskal-Wallis analysis in Chapter 3,

(iv) To use a versatile MS/MS fragmentation method from which the results can be compared to database spectra to identify structures of interest of ions previously determined to be statistically significant.

(v) To compare the compounds identified from Chew Stoke sewage works outfall to those previously observed in wastewater using targeted analyses in the literature.

As discussed in Section 3.1 sewage treatment works outfalls have been the subject of extensive targeted analysis programmes, which have identified a necessarily limited range of compounds. This chapter will consider two hypotheses: 1. *That many of the compounds determined in the DI-HRMS spectra presented in Chapter 3 to have been contributed to the river from the sewage outfall will be the same as those compounds identified in previous targeted studies e.g. known*

micropollutants, and 2. *That as a result of the untargeted approach adopted herein a number of anthropogenic compounds will be detected that have been overlooked in previous studies e.g. unknown micropollutants*.

4.3 Analysis of DOM extracts from Chew Stoke sewage treatment works using HPLC-MS

HPLC was performed as outlined in Section 2.3.3. Figure 4.1 shows the total ion chromatograms (TICs) obtained by HPLC-Orbitrap MS of Blank HPLC water, upstream, downstream and sewage outfall SPE extracts. All extracts showed a large peak at 1 min which likely originated from the HLB cartridge used for the SPE. Even though there are clear visual differences between the three TICs the extensive coelution makes meaningful comparison of the different sample extracts impossible. The poor separation shown in the TIC is due to the high complexity and chemodiversity of DOM in the SPE extracts. Figure 4.1(b) shows the TIC of the upstream SPE extract. At 10 min there are no distinct chromatographic peaks which could indicate the elution of a discrete compound. Furthermore, the mass spectrum shown in Figure 4.2(c) shows that there are numerous ions derived from compounds coeluting between 10 and 10.10 min. However, selecting ions from this complex spectrum and plotting their extracted ion chromatograms (EICs) shows two very different phenomena. The EIC of the ion *m/z* 616.3170 shown in Figure 4.2(a) shows a clearly eluting peak at 10.21 min. In contrast the EIC of *m/z* 381.1902 in Figure 4.2(b) shows a broad peak spanning 10.00 min of the chromatogram. This shows that this compound does not separate effectively with the C_{18} column. By using different columns or possibly two-dimensional liquid chromatography better resolution of this mixture may be possible. However, because of the complexity of these extracts it is likely that there will always be a significant number of compounds which do not resolve well using different columns.

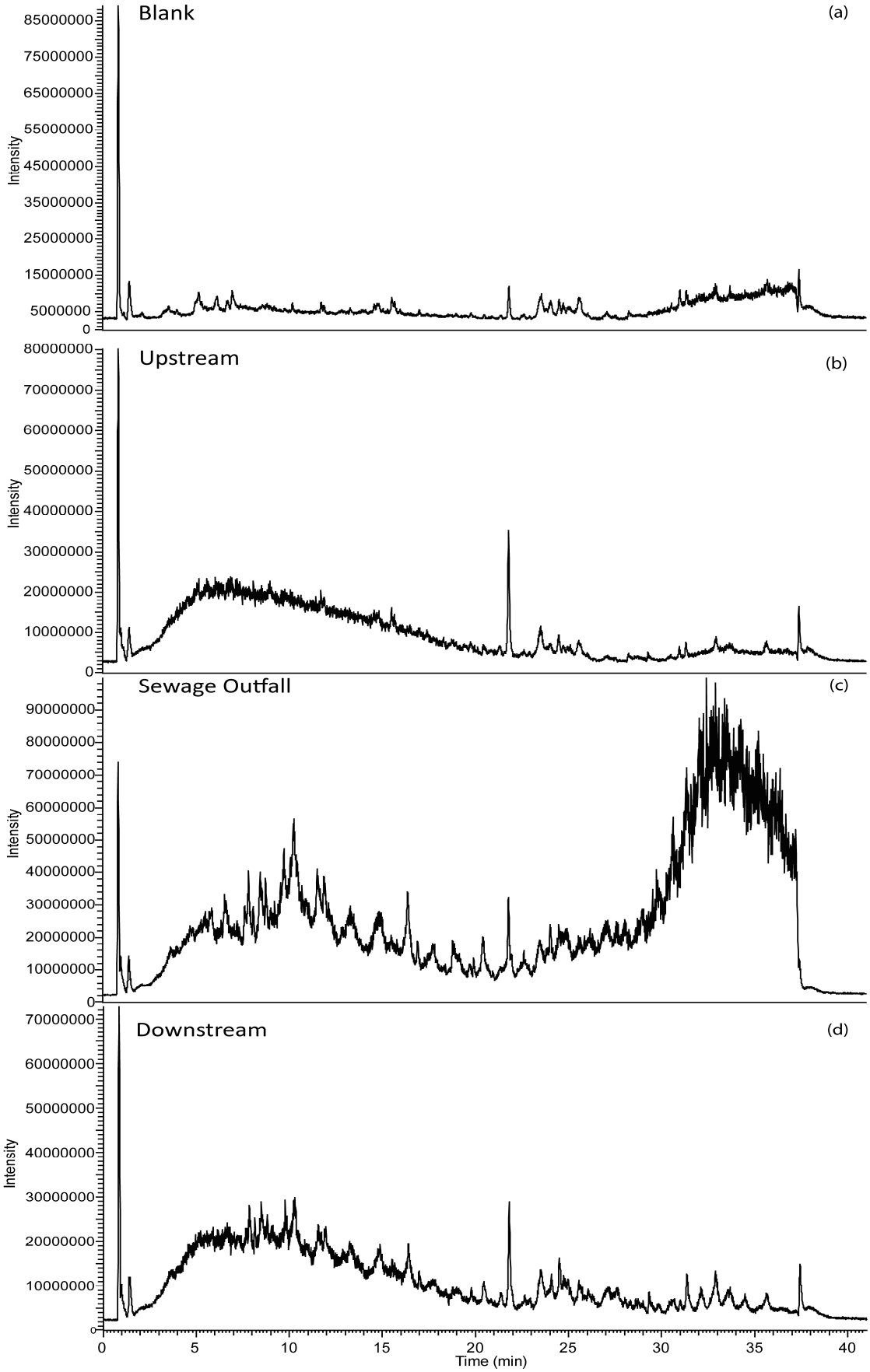


Figure 4.1. TIC of the HPLC-MS data collected comparing: (a) Blank SPE water extract, (b) upstream river water SPE extract, (c) sewage outfall water SPE extract, and (d) downstream river water SPE extract.

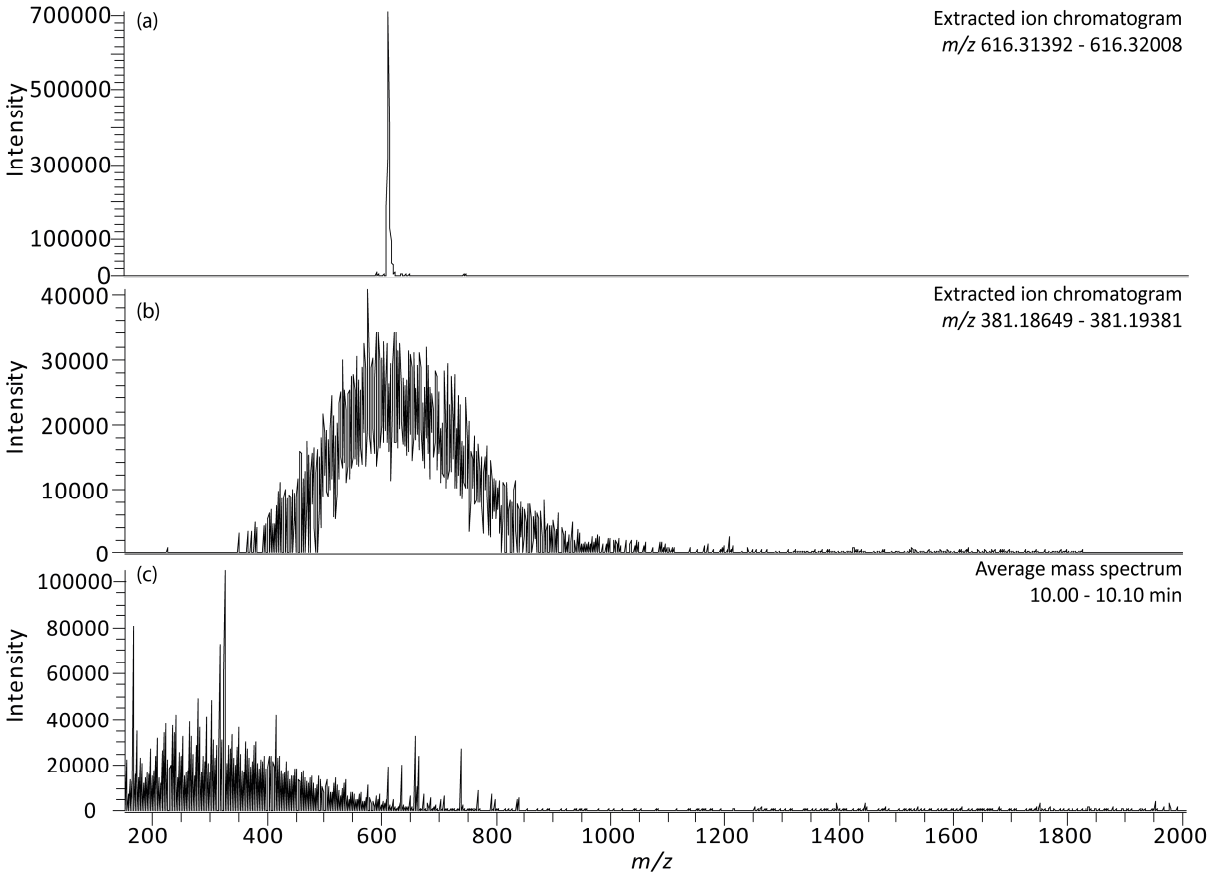


Figure 4.2. HPLC-MS analysis of the upstream SPE extract. (a) EIC of ion *m/z* 616.3170 using a 10 ppm mass range, (b) EIC of ion *m/z* 381.1902 using a 10 ppm mass range, and (c) average mass spectrum of ions eluting between 10.00 to 10.10 min for the range *m/z* 150 to 2000.

Figure 4.3 shows the chromatograms for absorbance at 254 nm. This wavelength was selected as it has been correlated to the concentration of DOM in rivers and is used by *in situ* sensors (Weishaar *et al.*, 2003). The chromatograms show an absorbance peak corresponding to the peak seen in the TIC at 1 min in all the extracts. In the blank extract there are no other significant features in the UV absorbance chromatogram apart from the gradual increase of absorbance as the concentration of acetonitrile is increased. However, in the downstream, upstream and sewage outfall extracts there is a second broad peak next to the peak at 1 min. Furthermore, there is an increase in the absorbance spectra between 3 and 6 min which then gradually reduces towards 20 min. This coincides with the broad hump of unresolved compounds in the TIC between 3 and 20 min common to all three extracts. This indicates that the compounds eluting in this region of the chromatogram are likely to be part of the fraction measured by *in situ* sensors and characterisation of this material may provide insights into the composition of the organic matter measured using sensor methods.

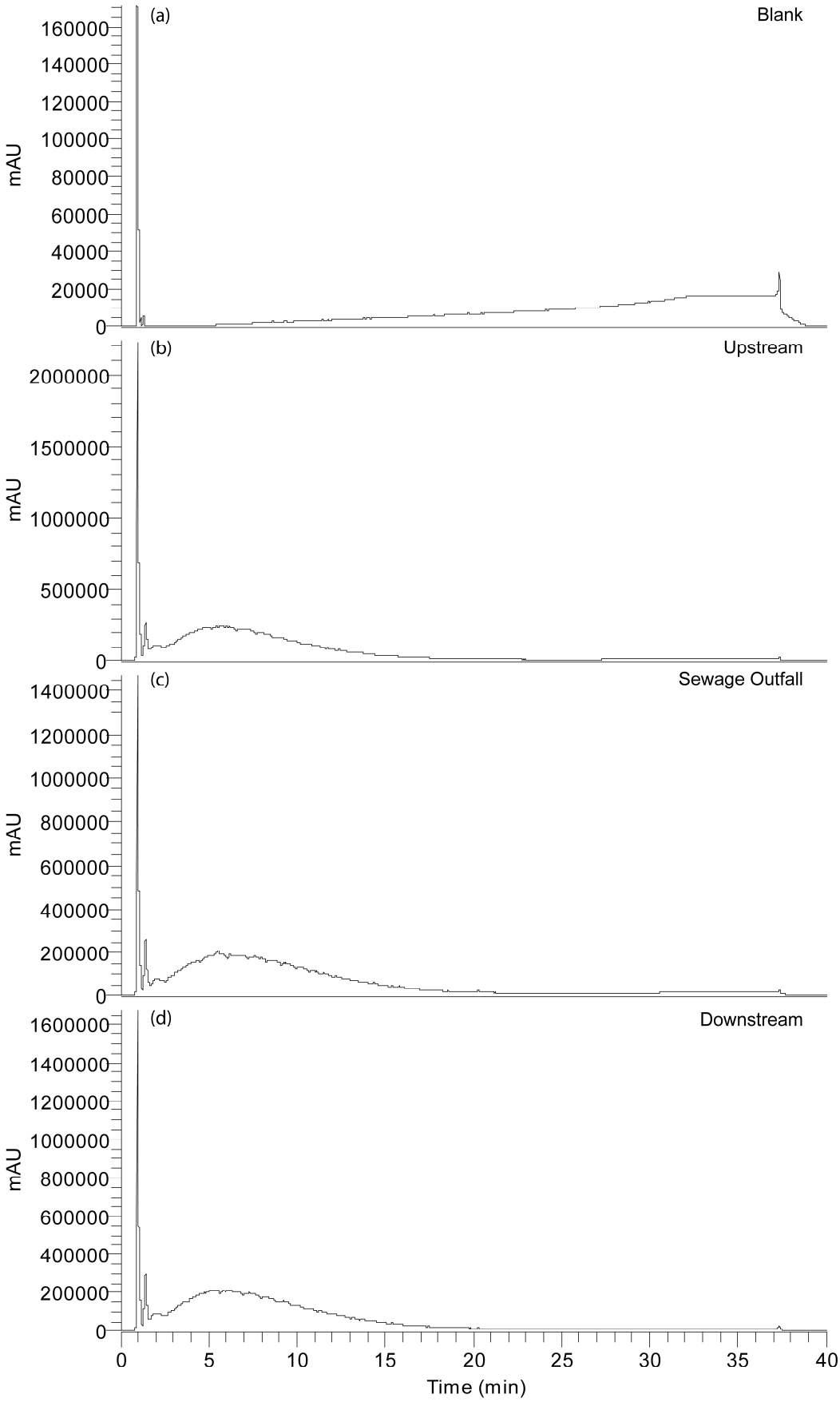


Figure 4.3. Specific UV absorbance HPLC chromatograms at 254 nm of the (a) blank, (b) upstream, (c) sewage outfall, and (c) downstream SPE extracts

The ion *m/z* 260.1647 was found to have a p value of 0.0013 when the sewage outfall and upstream DI-HRMS spectra were compared using Kruskal-Wallis analysis. Figure 4.4(a) shows the EIC of the ion *m/z* 260.1647 from the HPLC-MS of the sewage outfall DOM SPE extract which shows a clear chromatographic peak eluting at 8.82 min. Analysing the mass spectrum at this retention time in Figure 4.4(b), shows that there are clearly many different molecular ions indicating numerous coeluting compounds characterised by this ion of interest. However, looking at a 2 Da mass range around the ion of interest shown in Figure 4.4(c) the MS spectrum shows that there is sufficient separation for the Orbitrap to isolate this particular ion of interest and selectively fragment it to determine its identity.

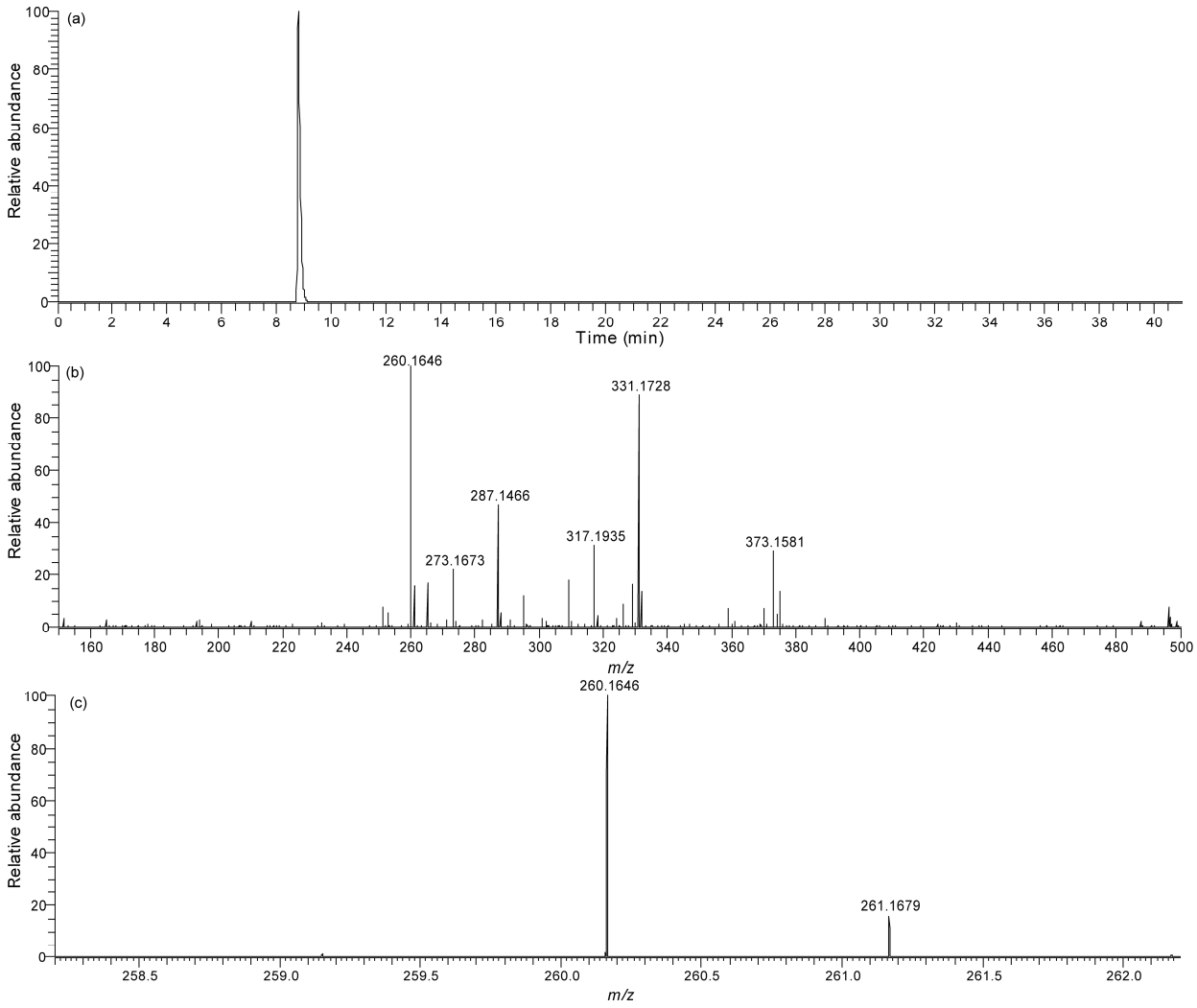


Figure 4.4. HPLC-MS analysis of Chew Stoke sewage outfall SPE extract (a) EIC of the ion *m/z* 260.1647 using a 10 ppm mass range *m/z* 260.1634 to 260.1360, (b) Mass spectrum at 8.78 min of the sewage outfall SPE extract for the mass range *m/z* 150 to 500, and (c) mass spectrum at 8.78 min of the sewage outfall SPE extract for the mass range *m/z* 258.2 to 262.2.

Figure 4.5 shows the EIC of the ion at *m/z* 260.1647 using a 10 ppm range *m/z* 260.1634 - 260.1360. Figure 4.5(a) and (b) show that there is no chromatographic peak detected in the blank or upstream DOM SPE extracts. The EIC in Figure 4.5(c) shows a large well resolved chromatographic peak in the sewage outfall which has a peak area of 9172857. Furthermore, the EIC Figure 4.5(d) shows there is a chromatographic peak in the downstream DOM SPE extra which, has a peak area of 3565401. This shows that this compound is originating from the sewage outfall and remains detectable downstream. The downstream peak is 38.9 % smaller than the peak detected in the sewage outfall this is likely due to the dilution of the outfall in the river. It should be noted that it remains possible that the reduction of peak size maybe partially due to loss of this compounds, either from the river e.g. binding to sediment or uptake by an organism, or the transformation of this compound into other products with a different mass. Only by identification of this compound further testing could the this be fully determined.

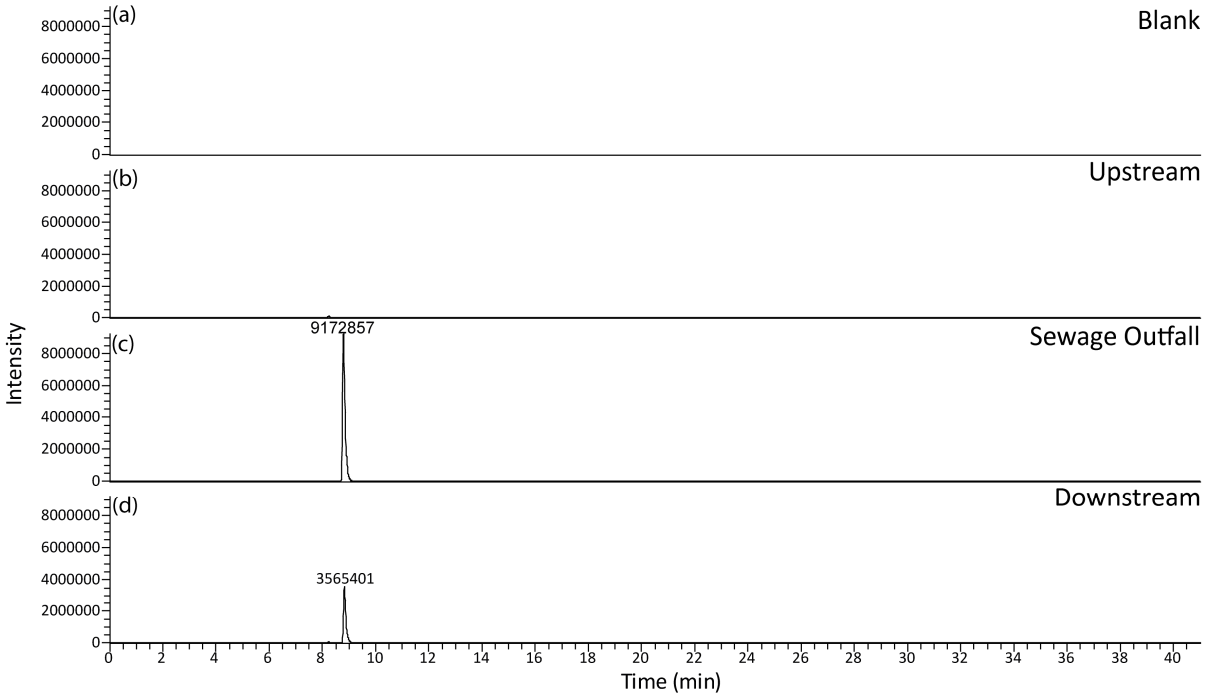


Figure 4.5. A comparison of the EIC for the ion *m/z* 260.1647 using a 10 ppm mass range *m/z* 260.1634-260.1360, from the HPLC-MS analyses of the (a) blank, (b) upstream, (c) sewage outfall, and (d) downstream SPE extracts. The peak area for the EIC peak in the sewage outfall and downstream spectra are displayed above the peak.

4.3.1 Peak picking and alignment

To further investigate the differences in the composition of each DOM SPE extract, peak picking and alignment were carried out for all HPLC-MS runs using the package XCMS with the centWave algorithm as described in Section 2.3.3.1. (Smith *et al.*, 2006; Tautenhahn *et al.*,

2008; Benton *et al.*, 2010). This created a data matrix of the peaks detected in the HPLC-MS analyses aligned by exact mass and retention time of an ion. In total 17,494 peaks were detected across all SPE extracts, again emphasising the complexity of riverine DOM. All peaks detected in the blank extract were removed leaving 14,325 peaks with retention times and masses unique to the river.

The peak areas of each ion with the same mass and retention time were compared for the different SPE extracts. The ratios of an ion's peak area were calculated as a percentage by comparing the individual upstream, sewage outfall and downstream SPE extracts to the sum of all extracts. The ratios of the peak areas were compared using a ternary plot. Figure 4.6 shows a ternary plot of the 14,325 ions detected and aligned across the three DOM extracts. Three sections of Figure 4.6 are highlighted showing different correlations between the extracts:

- Green highlights ions where less than 5% of the total peak area is found in the upstream extract.
- Blue highlights ions where less than 5% of the total peak area is found in the sewage outfall.
- Red highlights ions where more than 5% of the peak area is found in the all 3 DOM extracts.

Most of the peaks (11,308) plot in the green section of the ternary plot. This shows that these peaks have a higher peak area in the sewage outfall and downstream extracts and lower peak area in the upstream extracts. The molecular ions relating to these peaks are therefore likely to originate from the sewage outfall. The peaks in the green area predominantly plot between 100 % and 50 % along the sewage outfall axis showing that these ions have a larger peak area in the sewage outfall than the downstream SPE extracts; as was expected. As shown in Figure 4.5 compounds originating from the sewage outfall will be diluted in the river so, the peak area in the downstream will be less than that found in the sewage outfall.

The red area of the plot highlights 2,593 ions in the centre of the plot showing that peak area of these ions is similar across all three SPE extracts. This shows that these molecular ions are found in both the river and the sewage outfall. Therefore, part of the composition is common to all three extracts.

367 ions plot along the downstream axis highlighted in blue which shows that when comparing the ratio of the peak areas across the SPE extracts peak area < 5% from the sewage outfall. The

ions plot towards the centre of downstream axis which shows that these have a similar peak area in the downstream and upstream DOM extracts. This shows that these ions are found in the river but not in the point source.

None of the ions plot along the upstream axis. This would be indicative of compounds found in the upstream and sewage outfall SPE extracts, which were not found in the downstream SPE extracts. Furthermore, few molecular ions plot near 100 % in the downstream extract which would be indicative of ions found only in the downstream extract. This would not be expected, as the river water DOM found downstream should be an admixture of the sewage outfall and upstream water.

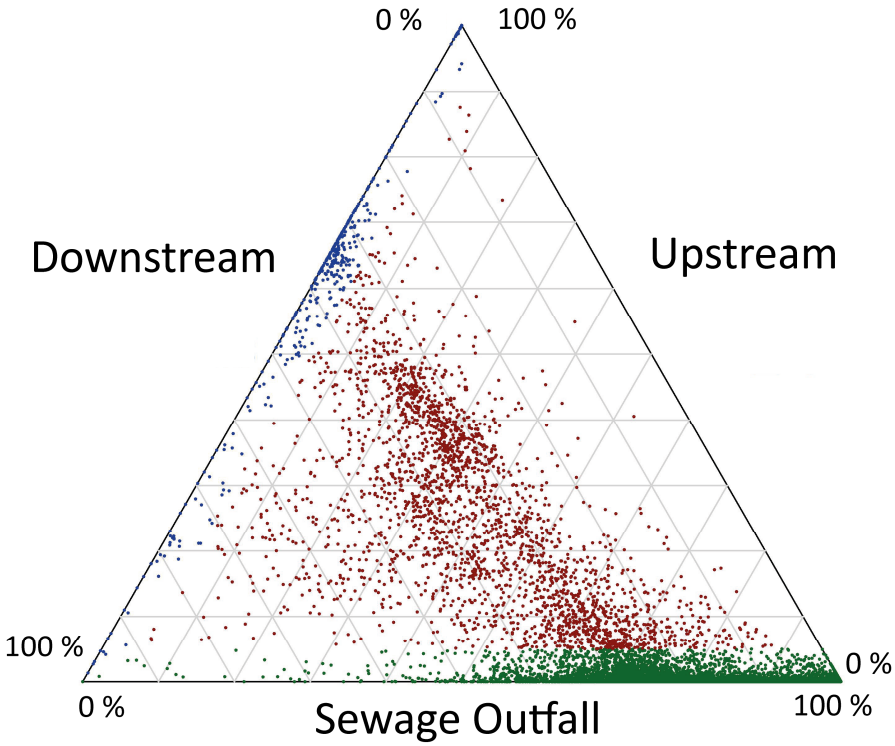


Figure 4.6. Ternary plot of ratios of the peak areas of the upstream, downstream and sewage outfall for all ions detected in the HPLC-MS analysis. < 5 % in the sewage outfall (blue), < 5 % in the upstream (green), > 5% in all samples (red)

The Kruskal-Wallis analysis of the DI-HRMS spectra detailed in Chapter 3, revealed 510 ions with a p value < 0.005, which had a higher abundance in the sewage outfall DI-HRMS spectra when compared to the upstream DI-HRMS spectra of the SPE extracts. These were chosen as statistically significant ions which could be compounds originating from the sewage outfall. The 510 ions were compared to the masses of ions detected in the HPLC-MS analysis. Ions found in the HPLC-MS analysis within 10 ppm of the exact mass of the 510 ions were selected

for further analysis. 90 of 510 ions were not found in the HPLC-MS. A possible explanation as to why 90 of the ions may not have been detected in the HPLC-MS analysis is that these compounds may have remained bound to the column and not have eluted in under the HPLC conditions chosen. Using alternative HPLC columns or conditions may be possible to further investigate these compounds. The remaining 420 ions were found to correspond to 681 EIC peaks in the HPLC-MS analysis.

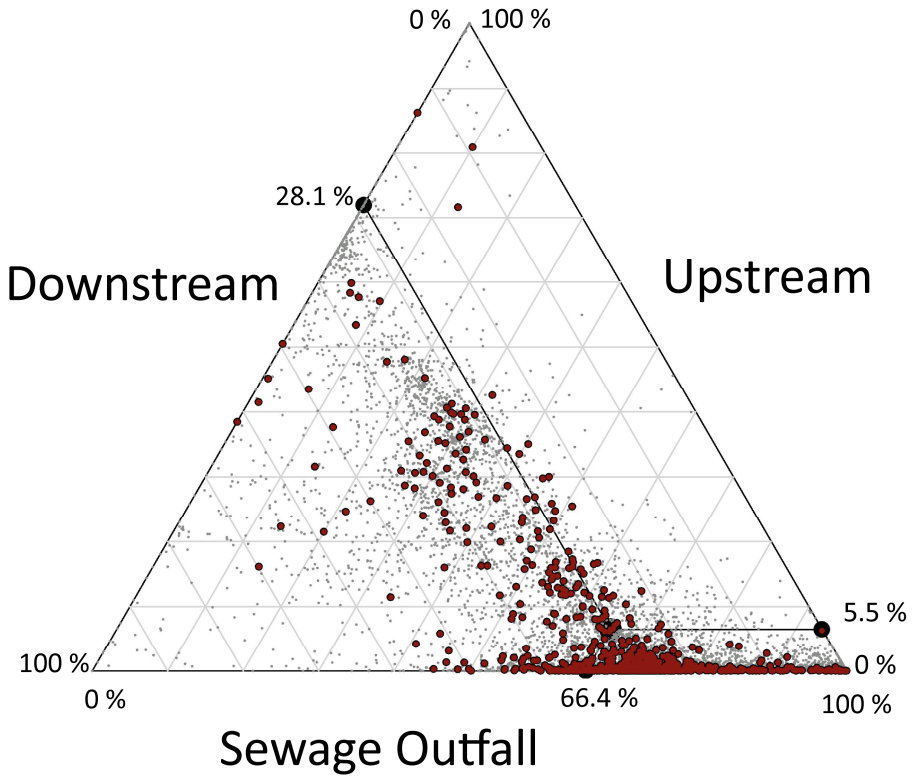


Figure 4.7. Ternary plot of ratios of the peak areas of the upstream, downstream and sewage outfall for all ions detected in the HPLC-MS analysis (grey), Ions with the same exact mass as the ions detected in the Kruskal-Wallis analysis are highlighted in red outlined in black. The mean percentage of the ions highlighted in red was calculated and plotted on each of the three axes.

Figure 4.7 shows a ternary plot of the ions detected in the HPLC-MS analysis. The 681 ions found to have the same exact mass as the 420 ions determined to be statistically significant from the Kruskal-Wallis analysis are highlighted in red. The mean of the percentage values of the 681 EIC peaks was calculated. The mean showed that overall the peaks selected by the Kruskal-Wallis analysis are strongly correlated with the sewage outfall with an average of 66.4 %. The average of the intensity values in the downstream was 28.1 % which is approximately half that of the sewage outfall which is due to the dilution of the effluent in the

river. 5.5 % was the average of the intensity found in the upstream, which is approximately a sixth of the downstream and a twelfth of the sewage outfall.

The ions highlighted in red predominantly plot along the sewage outfall axis which, as previously mentioned, shows a correlation between the composition of the sewage outfall and downstream SPE extracts. This shows that the peak areas of the ions highlighted in red have a low peak area or are not detected in the upstream extracts. This is also reflected in the mean. Furthermore, the peak areas of the ions plot between 50 % and 100 % along the sewage outfall axis showing that the peak area of an EICs are larger in the sewage outfall than the downstream. This is to be expected as a compound originating from a point source not found in the river would have a higher concentration in the point source water before being diluted in the river. Few ions plot towards the centre of the ternary plot indicating that these ions are common to all extracts. However, the exact mass of the ions of these peaks were found to have corresponding ions with the same exact mass but different retention times which were more intense in the sewage outfall.

In conclusion, the Kruskal-Wallis analysis has prioritised relevant ions which have a higher peak area in the sewage outfall than the upstream SPE extracts. These ions are less intense in the downstream extract showing the expected dilution effect of the river. However, there are clearly many other ions which plot in the area which would represent compounds coming from the sewage outfall which could be further investigated.

4.3.2 Compound identification using MS/MS.

Compounds can be identified from their characteristic product ion spectra using reference spectra. Hard ionisation used in MS generates ions by the fragmenting compounds. The product ions are characteristic of the compound being fragmented. Different hard ionisation methods will produce different fragmentation spectra. Although the ionisation source and fragmentation energy can be varied a set value is used for a single run. Similarly, there are standard energies which are commonly used across research studies. For example, for electron impact ionisation 70 eV is commonly applied. This standardisation is reflected in the database entries of reference spectra characterised using hard ionisation MS. Databases typically contain few reference spectra at commonly used fragmentation energies. However, these methods require good chromatographic separation of compounds prior to ionisation, otherwise this would result in an overlapping product ion spectrum.

Soft-ionisation ionises the compound by the addition of an adduct inducing little or no fragmentation (Dass, 2006). The molecular ion can then be measured before being isolated and fragmented. Different fragmentation methods can be used including CID, HCD, electron transfer induced dissociation and photodissociation (Dass, 2006). Each method will produce different fragments from the same molecular ion. Moreover, many of these methods can be used across different energies. Contemporary mass spectrometers such as the Orbitrap can use different fragmentation energies and fragmentation methods in a single run in the elution time of a chromatographic peak.

A result of this flexibility in soft ionisation MS means that many databases are not standardised. Reference spectra are recorded using different fragmentation methods, energies and instruments. Furthermore, databases sometimes only have one product ion spectra of a compound, thereby reducing the probability of getting an exact match. Some databases such as mzCloud require a reference spectrum to be recorded across multiple fragmentation energies and only accept spectra recorded on Orbitrap instruments, but this is not common practice. To maximise the possibility of matching product ion spectra to a database, a data dependant fragmentation method was designed to fragment the molecular ion through 6 fragmentation energies 10, 20, 30, 40, 50 and 60 eV using CID.

The 681 peaks in the sewage outfall SPE extract which correspond to the 510 ions highlighted as significant by the comparison of the DI-HRMS spectra using Kruskal-Wallis analysis were compiled into a target mass list and the retention times were determined from the HPLC-MS. MS^2 fragmentation was carried out as described in Section 2.3.4. In a single data dependant acquisition run, 1920 individual product ion spectra were recorded for 320 unique precursor ions.

The product ion spectra were searched through the databases mzCloud, which is a repository of spectra recorded using Orbitrap HRMS, and Massbank, which is the chosen database of the Norman Network which focuses on the analysis of environmental samples including water quality to enhance exchanges of information between different research organisations (Several *et al.*, 2009). Results from the database matches were filtered to include only high scoring matches, with relevant instrument type and fragmentation energy. These were then checked manually. The product ions spectra of 22 ions were exact matches to database recorded spectra. These are summarised in Table 4.1. This included a mixture of metabolites, pharmaceuticals, illicit drugs and flame retardants. Figure 4.8(a-c) shows the EICs for all the identified ions

(including PPG which is discussed in more detail below) from the HPLC-MS analysis of the upstream, downstream, sewage outfall SPE extracts which are normalised to the size of largest peak detected in the sewage outfall. Clearly none of the peaks found in the sewage outfall and downstream are identifiable in the upstream extract. All peaks are largest in the sewage outfall and smaller in the downstream extract. This is likely to show the dilution of the sewage outfall caused by mixing with the river. However, some of the reduction of peak size could also be due to the instream degradation or binding to particulate matter.

The EIC of the precursor ions of citalopram and carbamazepine-10,11-epoxide are shown in Figure 4.8 (d-g). For all peaks the peak is largest in the sewage outfall, smaller in the downstream and not detected in the upstream SPE extract. The EIC of Citalopram shows a single clearly resolved peak in the chromatogram. In comparison, for the EIC of carbamazepine-10,11-epoxide there are 3 clearly resolved peaks showing that there are multiple isomers with the same exact mass. This shows that a change in intensity of an ion in the DI-HRMS spectrum may be the collective change of multiple isomers.

The list of identified compounds includes 15 pharmaceuticals and pharmaceutical metabolites. The pharmaceuticals identified can be subdivided into their uses: pain management, mental health medication, heart medication, antiviral agents and antihistamines.

Two compounds were identified associated with pain management codeine and amitriptyline. Codeine is a commonly prescribed painkiller, often used to treat long term pain but also used for substance misuse (Baker *et al.*, 2014). Figure 4.9 shows the high level of similarity of the recorded product ion spectra to mzCloud reference spectra of codeine in both the accurate mass and abundances of product ions. It is a commonly identified compound in targeted sewage outfall studies (Stackelberg *et al.*, 2004; Loos *et al.*, 2012; Verlicchi *et al.*, 2012; Santos *et al.*, 2013). Amitriptyline is a medication used to treat nerve pain and depression. It has been the subject of targeted studies analysing both pharmaceuticals and antidepressants in sewage effluent and surface water (Fatta-Kassinos *et al.*, 2011; Lindholm-Lehto *et al.*, 2016). Furthermore, amitriptyline is metabolised by humans in the liver to the active compound nortriptyline. Nortriptyline is a prescribed antidepressant as well as a metabolite of amitriptyline and was also identified in the sewage outfall extracts. Like amitriptyline it has been detected in sewage treatment works and surface water. As nortriptyline is the major metabolite of amitriptyline it is not possible to determine whether nortriptyline is being used as a medication or is a metabolite of amitriptyline.

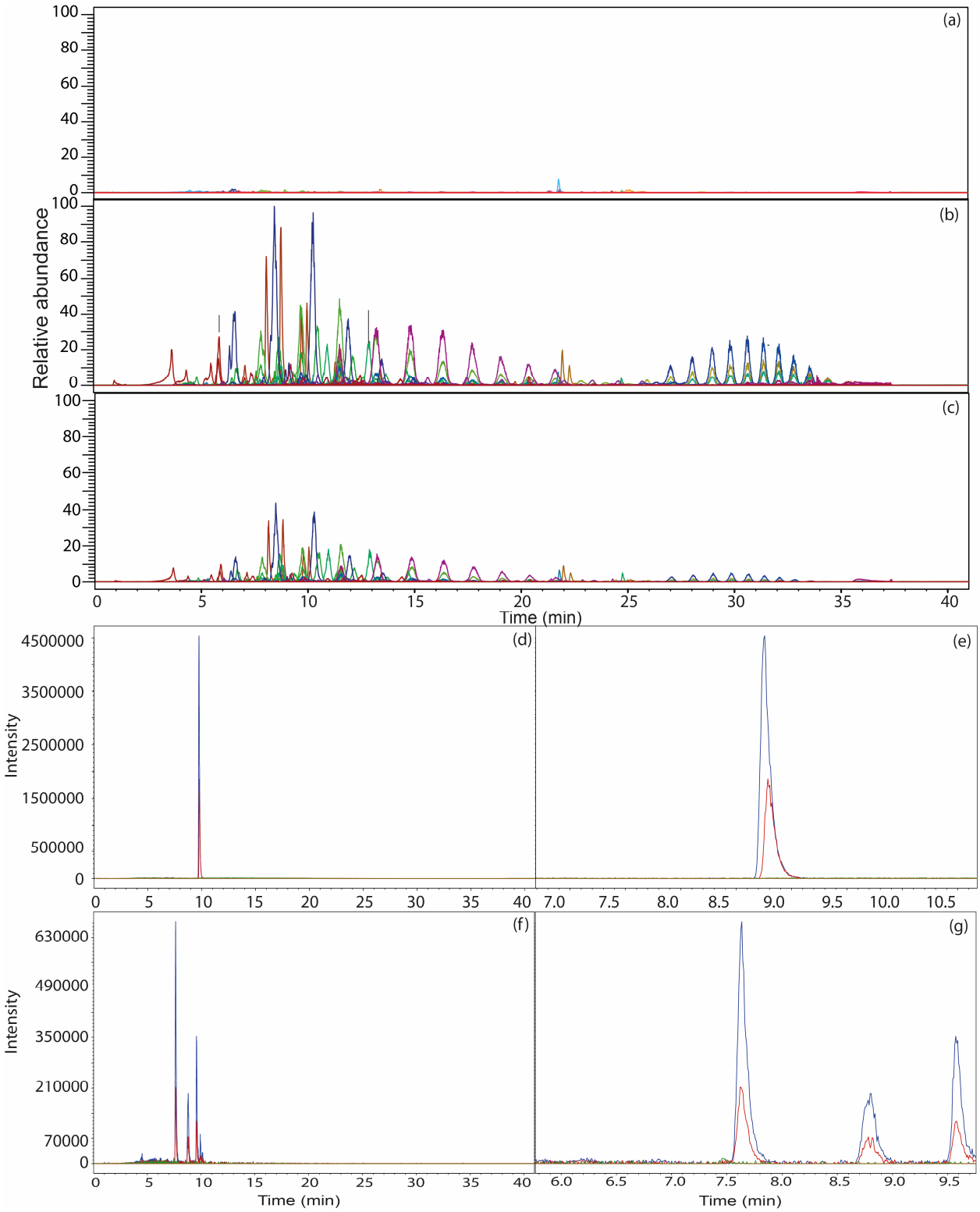


Figure 4.8. Overlaid EIC for all identified ions normalised to the highest abundance peak in the sewage effluent extract. from the HPLC-MS of the (a) upstream, (b) sewage outfall and (c) downstream extracts. Overlaid EIC of the sewage outfall (blue), downstream (red), upstream (green) and, blank (orange) SPE extracts for the ions: *m/z* 325.1678 to 325.1744 citalopram (d) 0 to 42 min and (e) 7.0 to 10.5 min, and *m/z* 253.0978 to 253.1003 carbamazepine-10,11-epoxide (f) 0 to 42 min and (g) 6.0 to 9.5 min

Table 4.1. Summary of the compounds identified in DOM SPE extracts using database searches of MS^2 spectra.

Precursor *(m/z)*	Retention Time (min)	Fragmentation energy CID (eV)	P value	Molecular Formula	Compound Match	Database	Product ions
300.1592	3.67	40	0.0013	$C_{18}H_{21}NO_3$	Codeine	mzCloud & Massbank	282.1489, 267.1254, 253.1242, 243.1033, 225.0910, 215.1067, 199.0736, 193.0648, 187.0754, 183.0804, 175.0754, 165.0699, 161.0597
268.1544	3.93	30	0.0016	$C_{14}H_{21}NO_4$	Atenolol acid	mzCloud	250.1441, 233.1176, 226.1079, 208.0971,191.0706, 165.0547, 145.0471, 116.1067, 98.0960
325.1915	5.47	30	0.0013	$C_{20}H_{24}N_2O_2$	Quinine	mzCloud & Massbank	307.1782, 279.1521, 278.1570, 264.1315, 253.1296, 226.1199, 210.0940, 202.0851, 198.0880, 186.0918, 184.0739, 174.0926, 172.0744, 166.1228, 160.0798, 134.0914, 110.0951
290.1392	5.89	30	0.0013	$C_{16}H_{19}NO_4$	Benzoylecgonine	mzCloud	272.1281, 168.1019, 150.0913, 124.11, 122.0968, 119.0493, 91.0544
256.0152	5.9	40	0.0013	$C_9H_7Cl_2N_5$	Lamotrigine	mzCloud & Massbank	229.0046, 221.0468, 220.0389, 213.99254, 210.98271, 193.04035, 186.9827, 185.9874, 183.9717, 179.0243, 173.98722, 171.97158, 166.0292, 165.02139, 158.9762, 151.0183
266.1657	6.07	40	0.0013	$C_{17}H_{19}N_3$	Mirtazapine	mzCloud	235.1230, 223.1230, 209.1073, 195.0917
304.1549	7.04	40	0.0013	$C_{17}H_{21}NO_4$	Cocaine	mzCloud	272.1281, 182.1176, 150.0913, 108.0807
253.0978	7.76	30	0.0013	$C_{15}H_{12}N_2O_2$	Carbamazepine 10,11-epoxide	mzCloud	254.0817, 236.071, 210.09187,180.0809
278.2113	8.14	30	0.0013	$C_{17}H_{27}NO_2$	Venlafaxine	mzCloud & Massbank	261.206, 215.1435, 121.0641
373.1586	8.75	30	0.0013	$C_{20}H_{24}N_2O_3S$	Desacetyl diltiazem	mzCloud	373.1580, 328.1002, 223.0900, 178.0321, 150.04
260.1647	8.82	30	0.0013	$C_{16}H_{21}NO_2$	Propranolol	mzCloud & Massbank	242.1539, 218.1176, 183.0804, 157.0648, 132.1019, 116.1070, 98.0943, 86.0964
325.1711	9.68	30	0.0024	$C_{20}H_{21}FN_2O$	Citalopram	mzCloud	325.1720, 307.16, 280.11, 262.10, 234.07, 166.07, 156.08, 116.05, 109.04
415.1456	10.06	30	0.0013	$C_{17}H_{20}F_6N_2O_3$	Flecainide	mzCloud	415.1454, 398.1189, 386.12,370.0870, 332.1345, 330.05569,318.0558, 315.1075 301.0297,
264.1752	11.15	30	0.0013	$C_{19}H_{21}N$	Nortriptyline	mzCloud	264.08, 233.13, 191.09, 155.09, 117.07, 105.07, 91.05
278.1909	11.29	30	0.0013	$C_{20}H_{23}N$	Amitriptyline	mzCloud	278.19118, 233.1328, 191.09, 179.09, 155.09, 117.07, 105.07, 91.05
502.2957	11.53	30	0.0013	$C_{32}H_{39}NO_4$	Fexofenadine	mzCloud	484.2830, 466.2726,262.1591, 250.5923, 246.1489, 233.1174,171.1168
237.1028	11.61	30	0.0013	$C_{15}H_{12}N_2O$	Carbamazepine	mzCloud	237.07, 220.08, 194.10, 192.08
192.1388	13.54	30	0.0041	$C_{12}H_{17}NO$	*N,N*-Diethyl-3-methylbenzamide, (DEET)	mzCloud	192.13829, 119.05, 100.07569, 91.0542

445.1636	13.21	30	0.0013	$C_{20}H_{21}FN_6O_5$	Raltegravir	mzCloud	361.1307, 318.1248, 278.0935, 253.0931, 236.0667, 193.08
286.1443	16.63	30	0.0013	$C_{17}H_{19}NO_3$	Piperine	mzCloud	287.1490, 201.0556, 173.06, 135.0447, 112.0756
399.2512	21.99	40	0.0013	$C_{18}H_{39}O_7P$	Tri(butoxyethyl) phosphate	Massbank	299.1627, 243.1001, 225.0894, 199.0736, 143.0108, 124.0100, 101.0963, 98.9841
273.1855	22.32	30	0.0013	$C_{18}H_{24}O_2$	Galaxolidone	mzCloud & Massbank	255.1743, 227.1794, 203.107, 175.1117

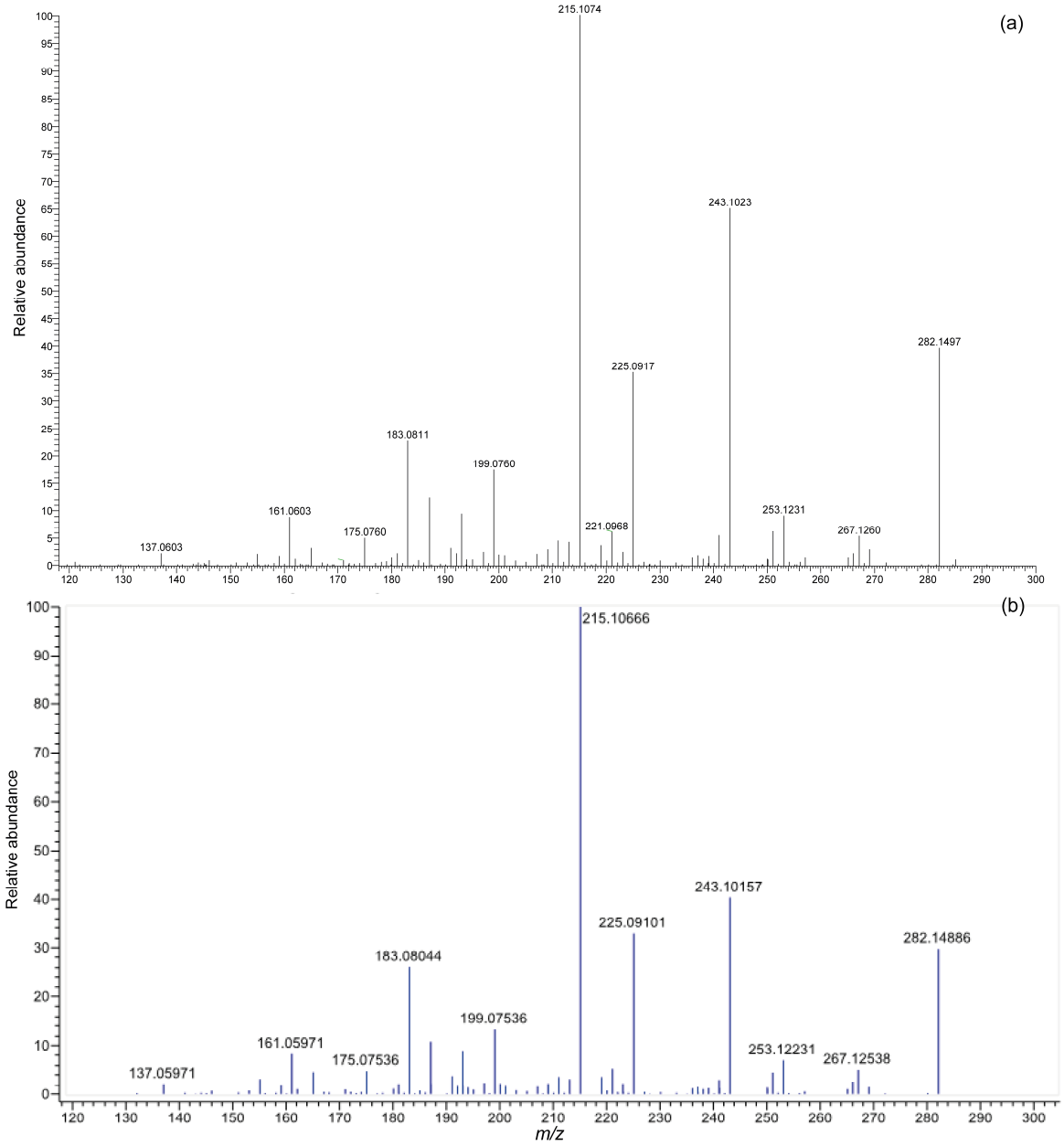


Figure 4.9. (a) Recorded MS^2 product ion spectra of codeine precursor ion *m/z* 300.15942 $[M+H]^+$, retention time 3.63 min CID 40 eV from the sewage outfall SPE extract and (b) mzCloud MS^2 product ion spectra of codeine precursor ion *m/z* 300.15942 $[M+H]^+$, CID 35 eV.

Four compounds identified are associated with heart medication. Propranolol and flecainide are active components of beta blockers and antiarrhythmic medication, respectively. Atenolol acid and desacetyl diltiazem are metabolites of the active ingredient atenolol, a beta blocker, and diltiazem, a calcium channel blocker. These are used in the treatment of hypertension and angina. Propranolol and atenolol are in the same class of beta blockers. In the metabolism of atenolol, the amide group undergoes hydrolysis to the carboxylic acid to form the metabolite atenolol acid. Even though atenolol is known to degrade and metabolise in this way, many targeted studies will only screen for atenolol, therefore potentially underestimating the

contribution of this pharmaceutical to surface waters, which could still potentially be biologically active. (Verlicchi *et al.*, 2012; Jeong *et al.*, 2017). Flecainide is a commonly used antiarrhythmic medication. In an EU survey it was found to be present in 98% of 90 sewage treatment plants tested (Loos *et al.*, 2012).

By far the most commonly identified compounds in this study were antidepressants, including, mirtazapine, venlafaxine, citalopram, and nortriptyline. Citalopram and venlafaxine are serotonin reuptake inhibitors which have been found in multiple wastewater treatment plants (Henry *et al.*, 2004; Fong and Hoy, 2012). Both citalopram and venlafaxine are known to affect the biota living in rivers. Notably both compounds have been shown to affect the behaviour of snails causing their feet to detach from substrates which can be potentially fatal (Fong and Hoy, 2012). Citalopram was found to effect the reproduction of cladoceran *Ceriodaphnia*, reducing the number of neonates produced by each female (Henry *et al.*, 2004). Mirtazapine is an atypical antidepressant which effects the levels of noradrenergic and serotonergic activity, and wastewater treatment plants have been identified as a source of this compound in the environment (Sheng *et al.*, 2014). Nortriptyline as discussed above is possibly a metabolite of the drug amitriptyline or a consumed antidepressant and has been identified in previous sewage treatment studies (Kasprzyk-Hordern *et al.*, 2007; Baker *et al.*, 2014; Sheng *et al.*, 2014). Despite the apparent biological effect that antidepressants are having on the environment these studies have been criticised for reproducibility, small population numbers and therefore the overall ecological impact of antidepressants are unknown (Vasquez *et al.*, 2014; Ford and Fong, 2016).

Two anticonvulsive active ingredients and one metabolite were identified from their product ion spectra. These were carbamazepine, lamotrigine and carbamazepine-10,11-epoxide. Lamotrigine is an anticonvulsant drug used in the treatment of epilepsy and bipolar disorder. Lamotrigine contains a 2,3-dichlorinated phenyl ring which produces a characteristic isotope ion pattern in the MS^1 scan prior to fragmentation relating to the combination of ^{35}Cl, ^{37}Cl and ^{13}C shown in Figure 4.10(a). Figure 4.10 (b) and (c) shows the recorded product ion spectrum for lamotrigine from the sewage outfall extract compared to the reference spectrum from mzCloud. Lamotrigine has previously been found in surface water (Wood *et al.*, 2017). Despite lamotrigine's biological activity little is known about its environmental impact. Carbamazepine is commonly used to treat epilepsy and the major metabolite carbamazepine-10,11-epoxide is also biologically active (Obach, 2013). This has been shown to originate from sewage treatment works (Miao *et al.*, 2005; Gros *et al.*, 2006; Loos *et al.*, 2012; Petrie *et al.*, 2014). Despite the

lack of research into the impact of these compounds on the aquatic ecosystem, these could potentially affect the neurological function of animals as they inhibit sodium channels in the nervous system.

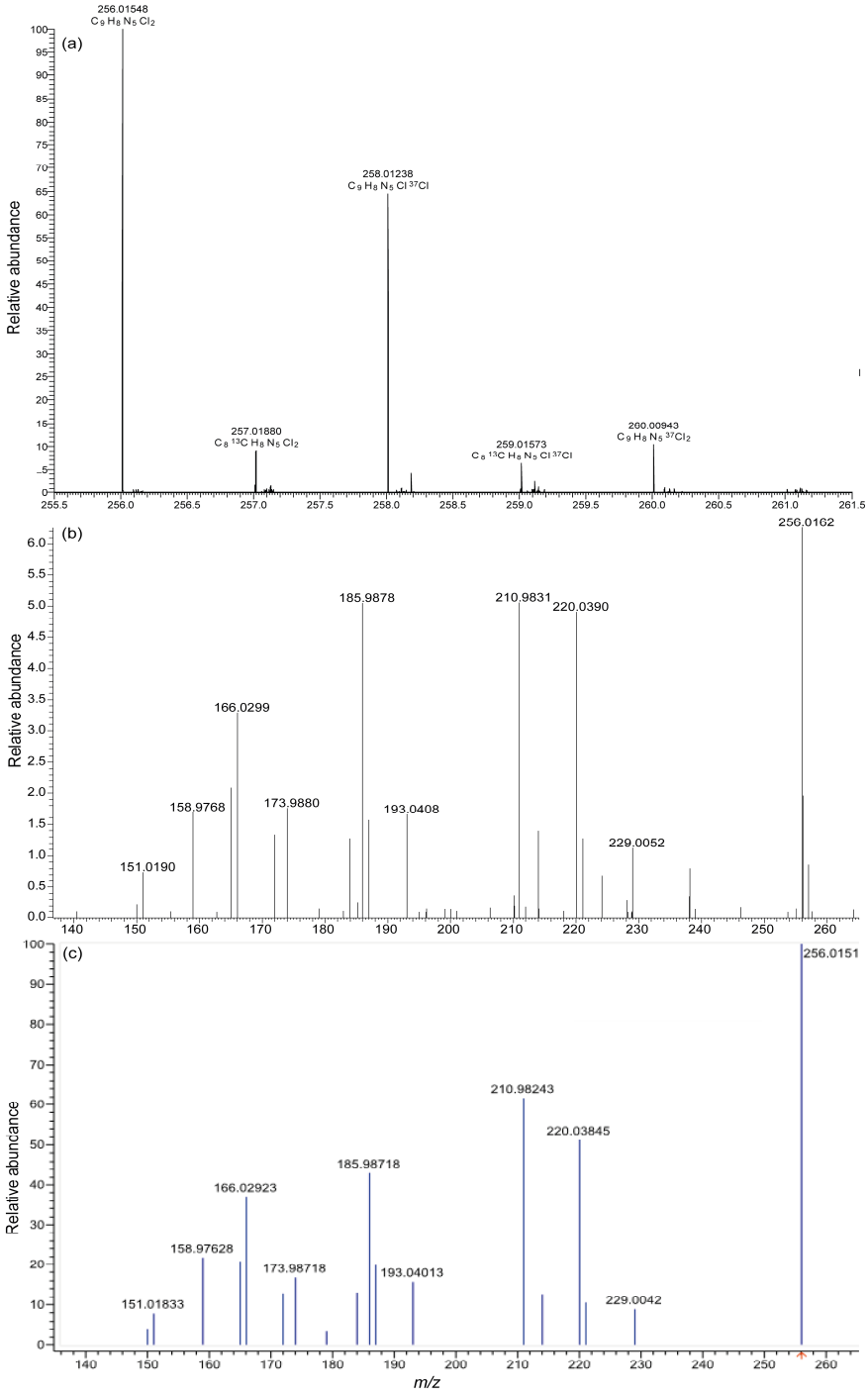


Figure 4.10. (a) Recorded MS^1 spectrum of lamotrigine $[M+H]^+$ at 5.84 min in the range *m/z* 255.5 and 261.5. Monoisotopic ion of lamotrigine *m/z* 258.01548 $C_9H_7N_5Cl_2$, ^{13}C isotopic ion of lamotrigine *m/z* 257.01880 $C_8{}^{13}C_1H_8N_5Cl_2$, ^{37}Cl isotopic ion of lamotrigine *m/z* 258.01238 $C_9H_8N_5Cl^{37}Cl$, ^{13}C ^{37}Cl isotopic ion of lamotrigine *m/z* 259.01573 $C_8{}^{13}C_1H_8N_5Cl^{37}Cl$, $^{37}Cl_2$ isotopic ion of lamotrigine *m/z* 260.00943 $C_9H_8N_5{}^{37}Cl_2$, (b) Recorded product ion spectra of lamotrigine precursor ion *m/z* 256.0154 $[M+H]^+$, retention time 5.90 min CID 30 eV from the sewage outfall SPE extract and, (b) mzCloud product ion spectra of lamotrigine precursor ion *m/z* 256.01513 $[M+H]^+$, CID 35 eV.

One antihistamine was detected in the sewage outfall samples fexofenadine. Unlike the other compounds it is a non-prescription based medication and targeted research studies focusing on antihistamines have previously identified this compound in sewage treatment works (Kosonen and Kronberg, 2009; Loos *et al.*, 2012). However, little is understood of its environmental impact.

One of the compounds identified as originating from the sewage outfall was Raltegravir, which is an antiretroviral drug used predominantly for the treatment of HIV and AIDs. This is also one of the three drugs prescribed by the NHS in post-exposure prophylaxis, a preventative treatment after potential exposure to HIV or AIDs. Previous targeted research focusing on surface water has not observed this before. This was unexpected as HIV is not prevalent in the UK, with only 101,200 people living with HIV, approximately 0.16 % of the population (AVERT, 2016). However, the prevalence of HIV may be underreported, or medication brought in from abroad. Therefore, the number of people registered as taking this drug is extremely low. mzCloud database and other studies have published product ion spectra of Raltegravir using MS to confirm its identity. The product ion spectrum correlates well with the observed fragmentation in Novak *et al.* (2010). To further confirm the identity of the compound, a data dependant acquisition was re-run targeting the *m/z* 445.1630 ion at 13.12 min under the same conditions as the database using HCD at 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 eV. All product ion masses and relative intensities matched the database. Two of the recorded product ion spectra are compared to two of the database spectra in Figure 4.11. However, a standard is needed to unequivocally identify and quantify this compound.

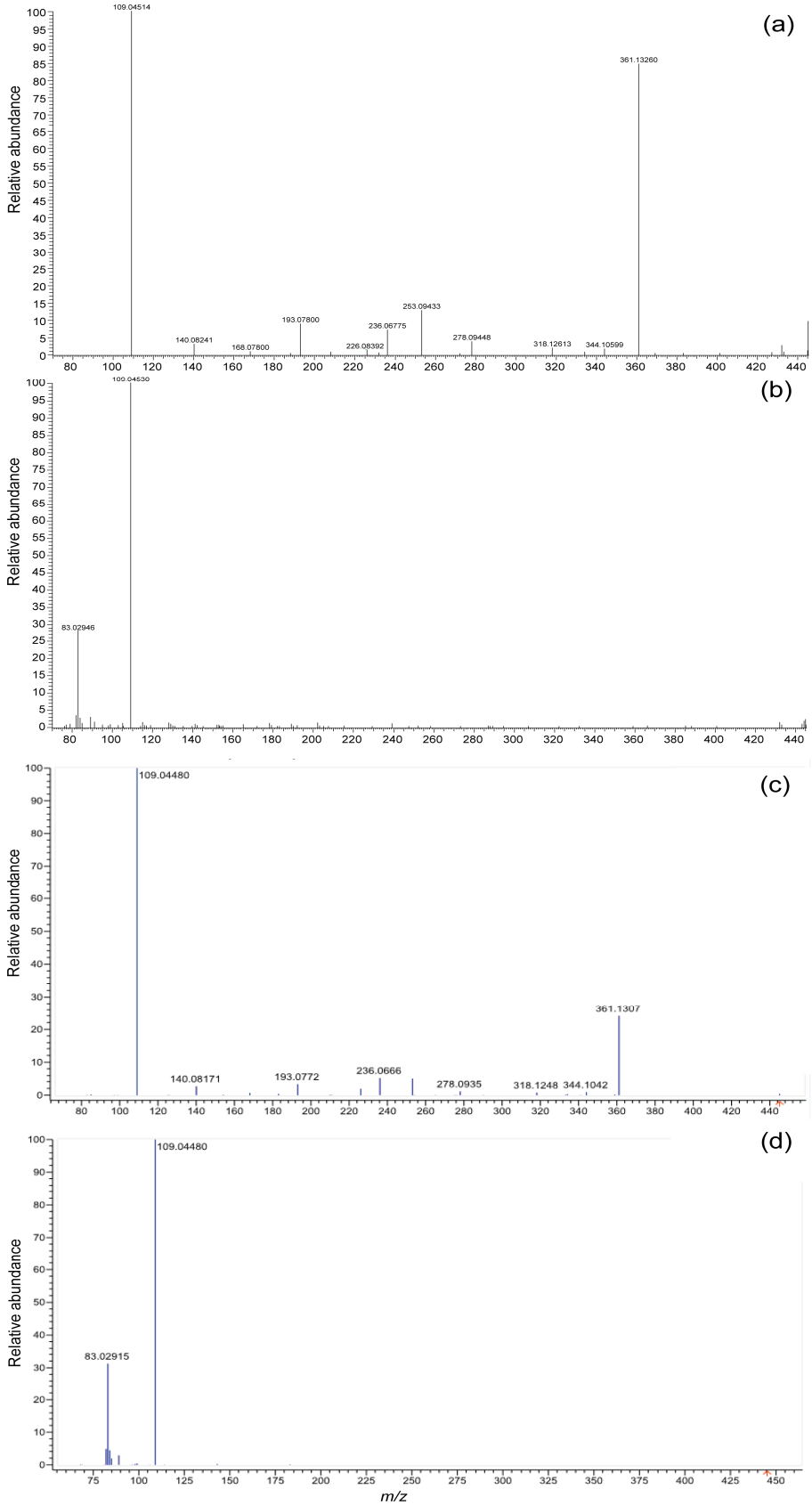


Figure 4.11. (a) Recorded product ion spectrum of Raltegravir $[M+H]^+$ precursor *m/z* 445.1603, retention time 13.21 min HCD 30 eV, (b) Recorded product ion spectrum of Raltegravir $[M+H]^+$ *m/z* 445.1603, retention time 13.21 min HCD 100 eV, (c) mzCloud product ion spectrum of Raltegravir $[M+H]^+$ *m/z* 445.1603, HCD 30 eV, (d) mzCloud product ion spectrum of Raltegravir $[M+H]^+$ *m/z* 445.1603 HCD 100 eV

Eight compounds not classed as pharmaceuticals were identified in addition to the 15 pharmaceuticals discussed above. The illicit drug cocaine was identified and its major metabolite benzoylecgonine, which is produced by the liver. These illicit drugs are routinely found in sewage outfalls and can be used to indicate the drug use in the surrounding area (Zuccato and Castiglioni, 2009). The quantification of illicit drugs and their metabolites have been previously used to estimate drug use in a population (Zuccato and Castiglioni, 2009; Baker *et al.*, 2014; Petrie *et al.*, 2014). The finding of illicit drugs particularly cocaine in surface water has been the subject of many media articles (Barkham, 2018). Cocaine and benzoylecgonine have been shown to effect the biota present in a river, in particular eels (Capaldo *et al.*, 2018). Eels are native to the river Chew where Bristol Water has a campaign to restock the eel populations with their conservation program "Spawn to Wild" (Bristol Water, 2014).

Two compounds were identified which can be attributed to natural food flavourings, quinine and piperine. Quinine is a naturally occurring compound commonly found in tonic water and has not been the focus of targeted studies. However, it is only present in the sewage outfall and downstream extracts, showing that it is not occurring naturally in the river. Quinine is also used as a preventative medication for malaria however, Britain is not an at-risk area for malaria and therefore it is unlikely that this is the use of this compound. However, when used for malaria prevention people take it for a long time; 6 weeks or more. Also, quinine sulphate is commonly used for night cramps and this may provide an alternative explanation why it is found. Therefore, this substance could be a potential biomarker for domestic sewage or a pharmaceutical. Piperine has also not previously been identified in sewage outfalls. The compound originates from black pepper and produces the pungent fragrance (Bajad *et al.*, 2003). Unlike anthropogenic compounds such as pharmaceuticals which are routinely identified due to their biological activity and known toxicity, quinine and piperine are both naturally occurring from plants. Both quinine which originates from the cinchona tree and black pepper from piper *nigrum* come from plants not native to the UK, but their products are routinely consumed in UK households. Therefore, the identification of flavouring's and organics associated with particular foods may allow inference as to the populations consumption habits and diet similar to the studies used to estimate a populations drug consumption from wastewater (Karolak *et al.*, 2010; Baker *et al.*, 2014).

DEET was identified in the sewage outfall extracts which is a commonly used insect repellent. It has been commonly identified in soils, wastewater influent and effluent (Loraine and Pettigrove, 2006; Murray *et al*., 2010). DEET has been subjected to environmental impact studies but has not been associated with any adverse environmental effects *(*Weeks *et al*., 2012).

The flame retardant and plasticiser tri(butoxyethyl) phosphate was identified in the sewage outfall. Flame retardants have been the subject of many targeted analysis studies including both brominated and phosphate containing compounds (Rodil *et al*., 2005). The environmental effects of phosphate containing flame retardants is currently unknown. However, inorganic phosphate has been shown to be a major driver in eutrophication hence the use of phosphorous stripping at sewage treatment plants, which includes Chew Stoke sewage treatment works. This shows that the phosphorous stripping of water does fully remove organic phosphate compounds, such as flame retardants.

Galaxolidone is a known metabolite of the fragrance galaxolide, which is commonly used in household cleaning products. This metabolite is known to be a product of biological degradation within sewage treatment works. It is a commonly identified compound in targeted studies (Franke *et al.*, 1999; Heberer, 2002).

The EICs in Figure 4.8 and ternary plot in Figure 4. 7 show that comparing the DI-HRMS spectra highlights compounds which are coming from the point source. The consistent differences in the peak area of compounds indicates that there is the potential to quantify these compounds in retrospect. However, this would require additional information including (i) calculation of the extraction efficiency of the compound using the HLB SPE cartridge, and (ii) quantification of the analyte by standard addition or using an isotopic standard. Using the compounds identified could provide a basis for more conventional targeted studies investigating, their concentration in other river systems or point sources and temporal changes in concentration.

The presence or concentration of a compound is not enough to determine if this compound will have an impact upon organisms within an aquatic ecosystem. As discussed above, most of the 22 compounds identified have a known biological activity in humans. However, there is a limited understanding of some of the compounds identified eco-toxicity in particular organisms. However, without identifying which compounds are present in the environment or within point sources their toxicity cannot be determined.

As discussed, all the 22 compounds identified have anthropogenic origins which explains why these would be found in a sewage treatment works. The ternary plots (Figure 4.6 & 4.7) highlight the numerous other components that remain unidentified. This shows the need to further investigate point sources using untargeted methods as known compounds can only account for a limited number of these components. Furthermore, more databases are needed to identify other compounds which record reference standards over multiple collision energies and mass spectral parameters to ensure that the database has as wide an applicability as possible.

4.3.2.1 Polypropylene glycol (PPG) identification

One hundred and seventy-seven ions were tentatively identified as PPG in the DI-HRMS spectra and summarised in Chapter 3 Section 3.4.4 Table 3.3 of which 157 had a p value < 0.005. The HPLC-MS analysis of the sewage outfall extracts showed that 16 out of the 18 PPG sequences were detected eluting from the column. The oligomeric ions from the PPG series 13 and 7 were not found in the HPLC-MS analysis. For all series the oligomers within their respective series separated on the column by molecular weight with lower molecular weight oligomers eluting earlier. The coelution of ions across different series was used to identify adducts and isotopes of the same compound.

Figure 4.12 shows the EIC for each ion in the PPG series 1, 6, and 9 from the HPLC-MS analysis of the sewage outfall SPE extract. The EIC shows a normal distribution of peak areas of the oligomers in series expected for a synthetic polymer. The same distribution is seen for each series of ions. The coelution of the oligomers within these three series and their molecular formulae show that series 1 is the sodiated adduct of PPG, and series 6 is the ^{13}C isotopic series corresponding to series 1, and series 9 is the ammonium adduct of the molecule.

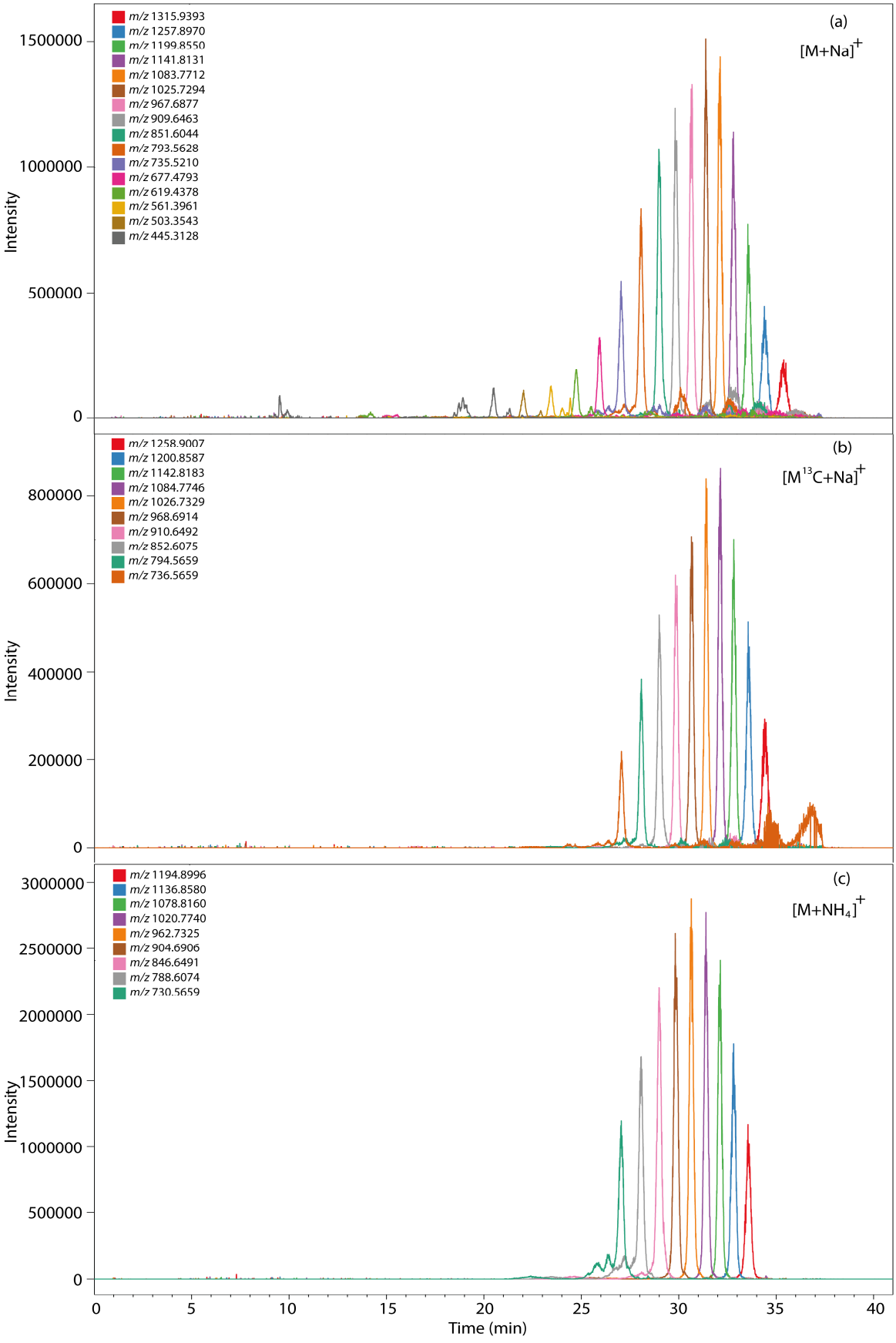


Figure 4.12. EICs from the HPLC-MS analysis of the DOM SPE extract of the sewage effluent. The ions extracted were identified to be oligomers of PPG from the DI-HRMS spectra (Table 3.3), the series displayed are: (a) PPG series 1 $[M+Na]^+$, (b) PPG series 6 $[M^{13}C+Na]^+$, and (c) PPG series 9 $[M+NH_4]^+$.

Figure 4.14 shows the recorded product ion spectrum of the *m/z* 1315.94 oligomer from PPG series 1, which contains four overlapping 58 Da spaced series of product ions. It is known from the literature that PPG produces a characteristic 58 Da spaced product ion spectrum (Okuno *et al.*, 2003). The fragments are derived from the elimination of $–(C_3H_6O)-$ monomer unit and H_2O from the parent molecule, suggesting that the ether bond undergoes a rearrangement to form a vinyl end group which is also observed in the CID of polyethylene glycol (Figure 4.16) (Chen *et al*., 2002).

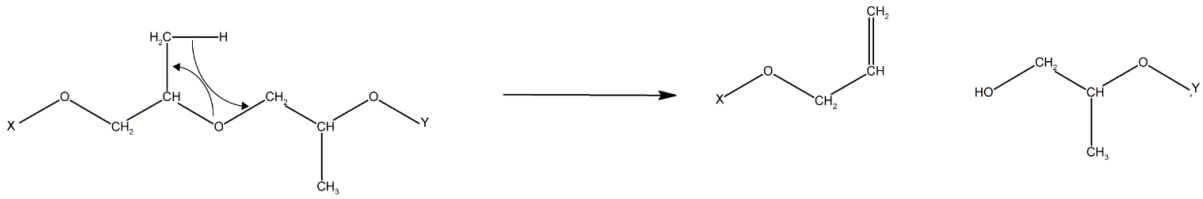


Figure 4.13. Proposed mechanism for the rearrangement during the fragmentation of the ether bond in PPG resulting in vinyl group (Okuno *et al.*, 2003).

Unlike the study by Okuno *et al*. (2003) which used the PPG diol for which 3 series of product ions were found, Figure 4.14 shows that there are clearly 4 series of 58 Da spaced product ions which indicates that the end groups are different as shown by the proposed fragmentation. Similar to a study by Chen *et al.* (2002) analysing the CID MS of PEG, which showed that the sodium adduct remained associated with the product ions, correlates with the product ion spectrum shown in Figure 4.14. From the product ion spectrum, it was possible to determine that the end groups of the polymer were hydroxyl and butyl end groups.

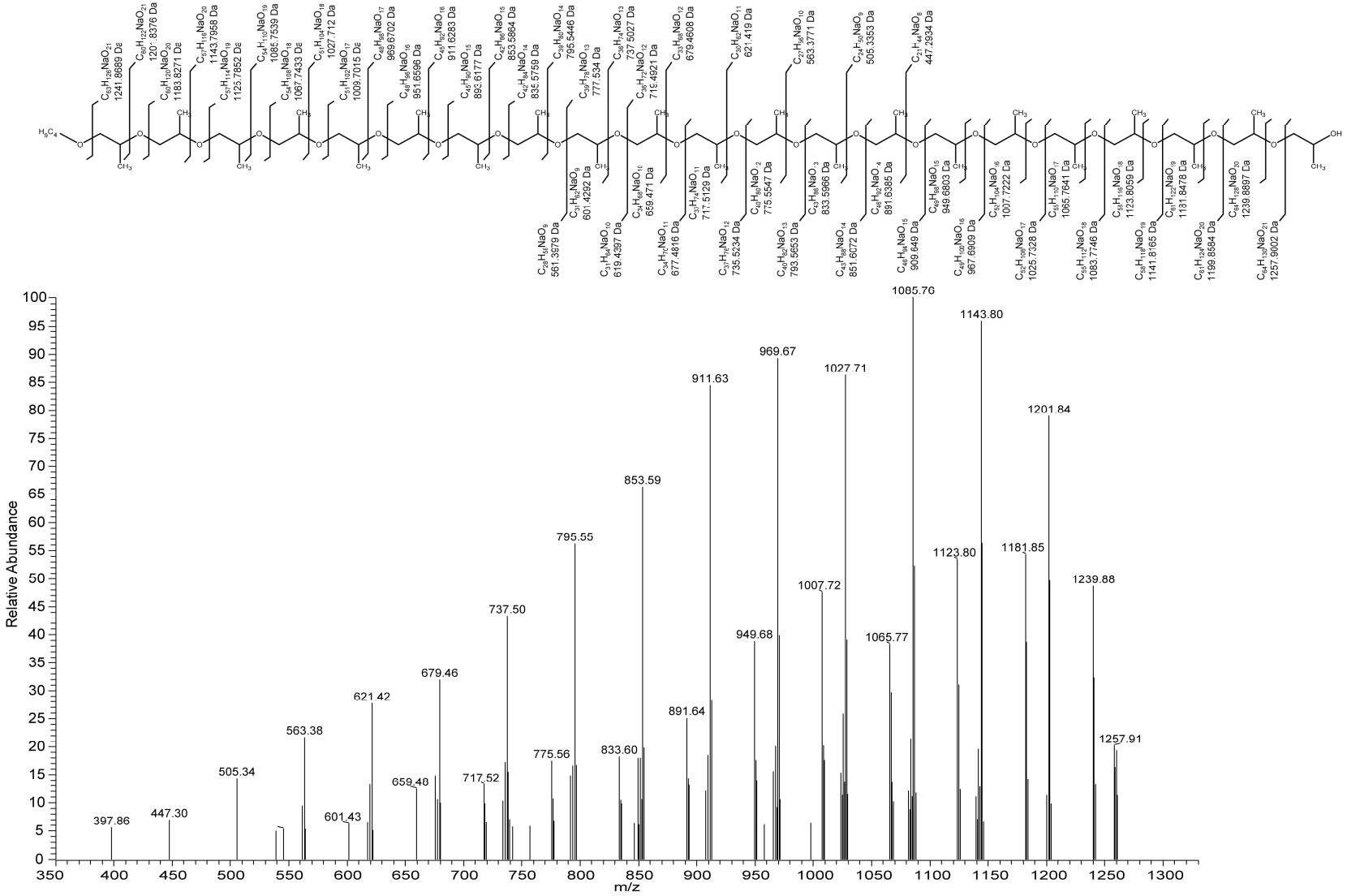
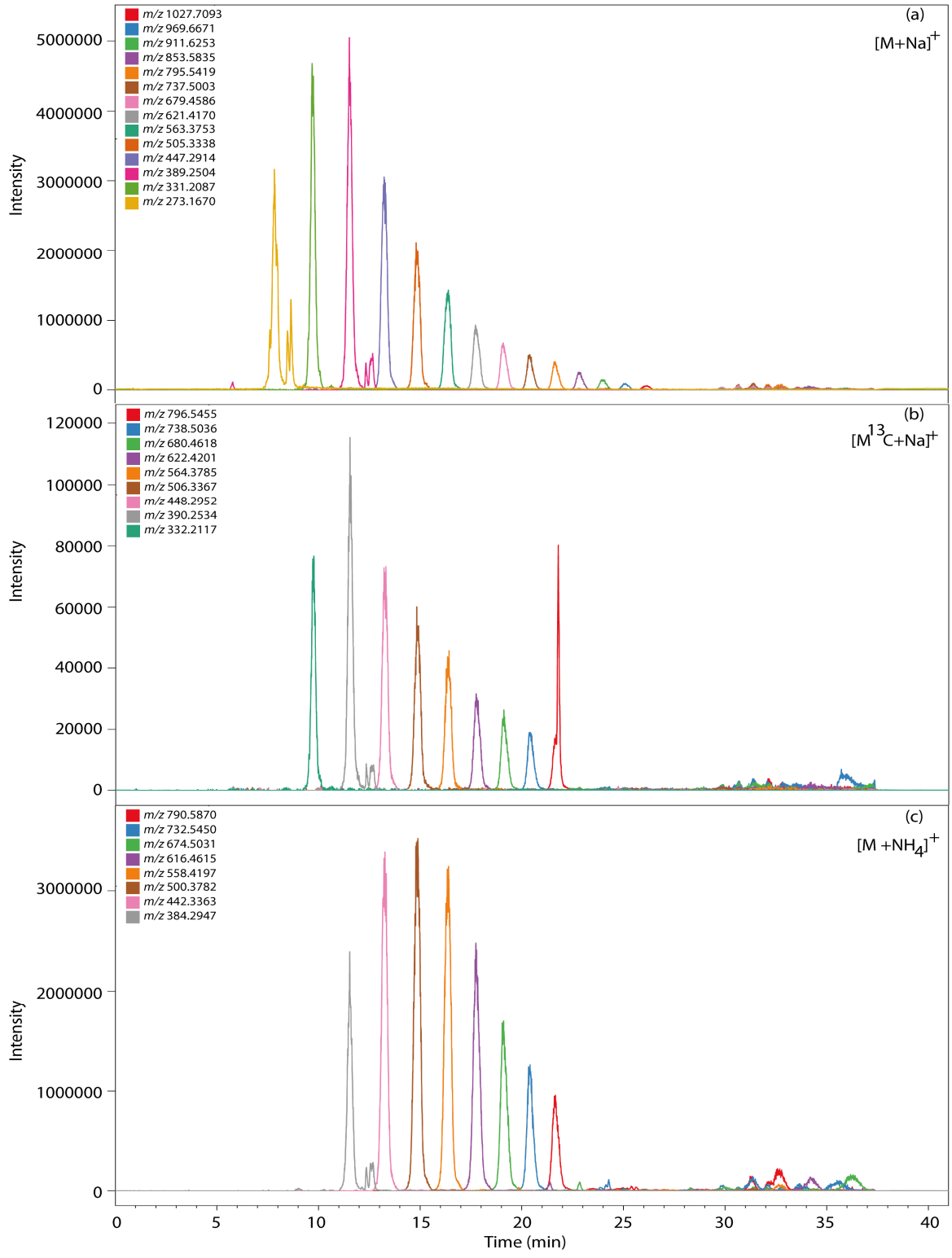


Figure 4.14. Proposed fragmentation pattern of the PPG molecule, associated with series 1 (top), for the recorded product ion spectrum of precursor ion *m/z* 1315.94 $[M+Na]^+$ 35.29 min CID 60 eV.

Figure 4.15 shows the EIC of the ions from PPG series 2, 3, 8, 15 and 16. From the EICs of the PPG series the 2, 3, 8 and 15 there are clear peaks corresponding to each oligomer in the polymeric series. However, the EICs of the ions from PPG series 16 shows multiple peaks for each oligomer. All ions in series 2, 3, 8 and 15 show a normal distribution of peak areas as expected for a polymer. The exception to this is the ion *m/z* 796.546 in PPG series 8 which does not fit the distribution of peaks. However, the peak shape is asymmetrical which suggests that there is another compound with the same exact mass coeluting with this oligomer.

From the molecular formulae it is possible to determine that the series 2, 3 and 15 are the same compounds but present as different adducts, namely: $[M+Na]^+$, $[M+H]^+$ and $[M+NH_4]^+$. Series 8 is the corresponding ^{13}C isotope series of PPG series 2. From the molecular formula series 16 would be expected to be the corresponding ^{13}C isotopes of series 3. There are EIC peaks for series 16 which correspond to the EIC of the oligomers in series 3, however, these are more difficult to determine due to the other peaks in the EICs of series 16.

Figure 4.16 shows the recorded product ion spectrum of an oligomer from PPG series 2. The spectrum shows are 2 series of product ions separated 58 Da. From the product ion spectrum, it is possible to determine that the end groups of PPG was the diol for which the proposed fragmentation of the molecule is shown above the spectrum in Figure 4.16.


(a)
[M+Na]+
m/z 1027.7093
m/z 969.6671
m/z 911.6253
m/z 853.5835
m/z 795.5419
m/z 737.5003
m/z 679.4586
m/z 621.4170
m/z 563.3753
m/z 505.3338
m/z 447.2914
m/z 389.2504
m/z 331.2087
m/z 273.1670
5000000
4000000
3000000
2000000
1000000
0
Intensity
(b)
[M13C+Na]+
m/z 796.5455
m/z 738.5036
m/z 680.4618
m/z 622.4201
m/z 564.3785
m/z 506.3367
m/z 448.2952
m/z 390.2534
m/z 332.2117
120000
100000
80000
60000
40000
20000
0
Intensity
(c)
[M +NH4]+
m/z 790.5870
m/z 732.5450
m/z 674.5031
m/z 616.4615
m/z 558.4197
m/z 500.3782
m/z 442.3363
m/z 384.2947
3000000
2000000
1000000
0
Intensity
0
5
10
15
20
25
30
35
40
Time (min)

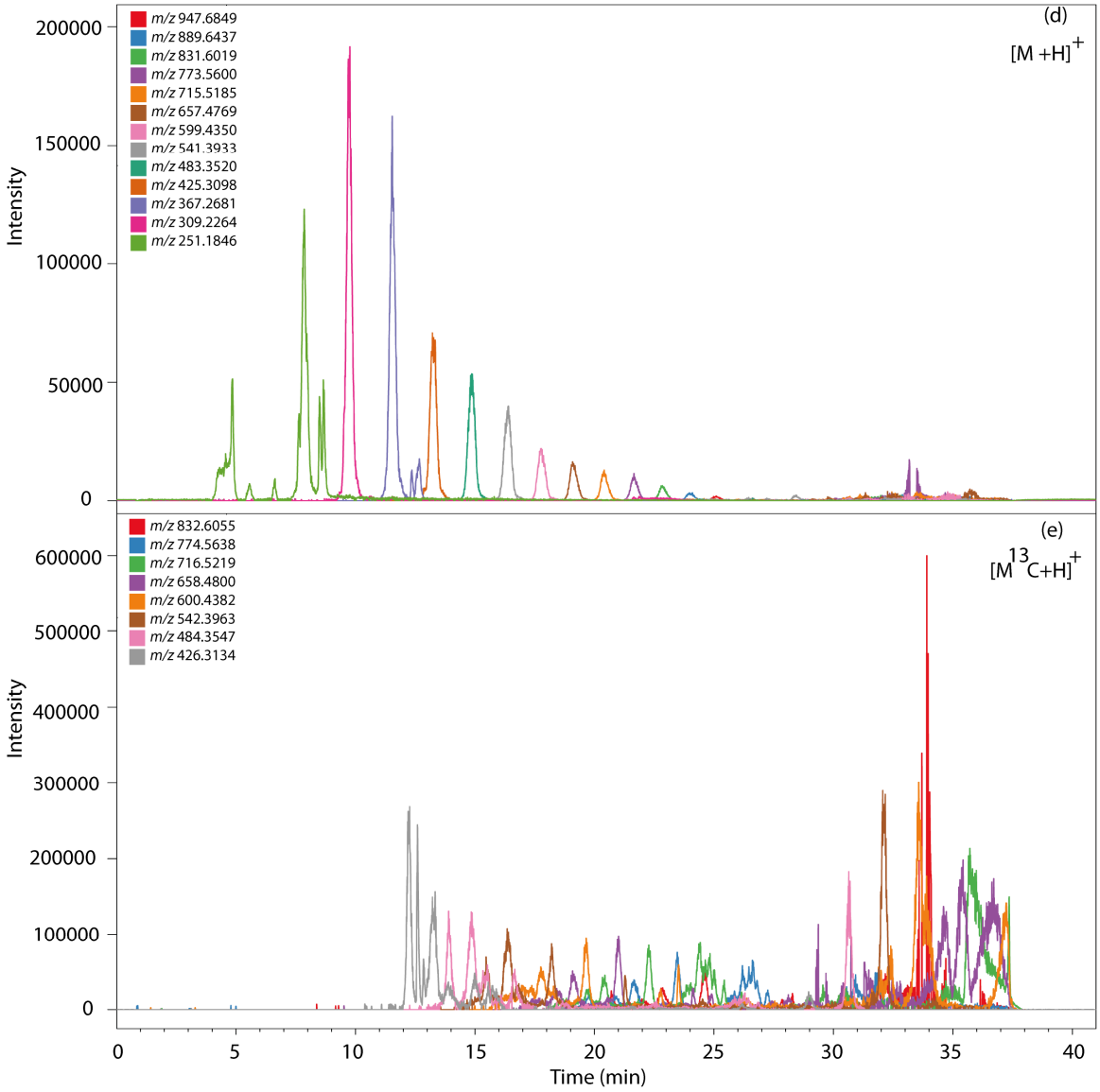


Figure 4.15. EICs from the HPLC-MS analysis of the DOM SPE extract of the sewage effluent. The ions were identified to be oligomers of PPG from the DI-HRMS spectra (Table 3.3), the series displayed are: (a) PPG series 2 $[M+Na]^+$, (b) PPG series 8 $[M^{13}C+Na]^+$, (c) PPG series 15 $[M^{13}C+Na]^+$ (d) PPG series 3 $[M+H]^+$, and (e) PPG series 16 $[M^{13}C+H]^+$

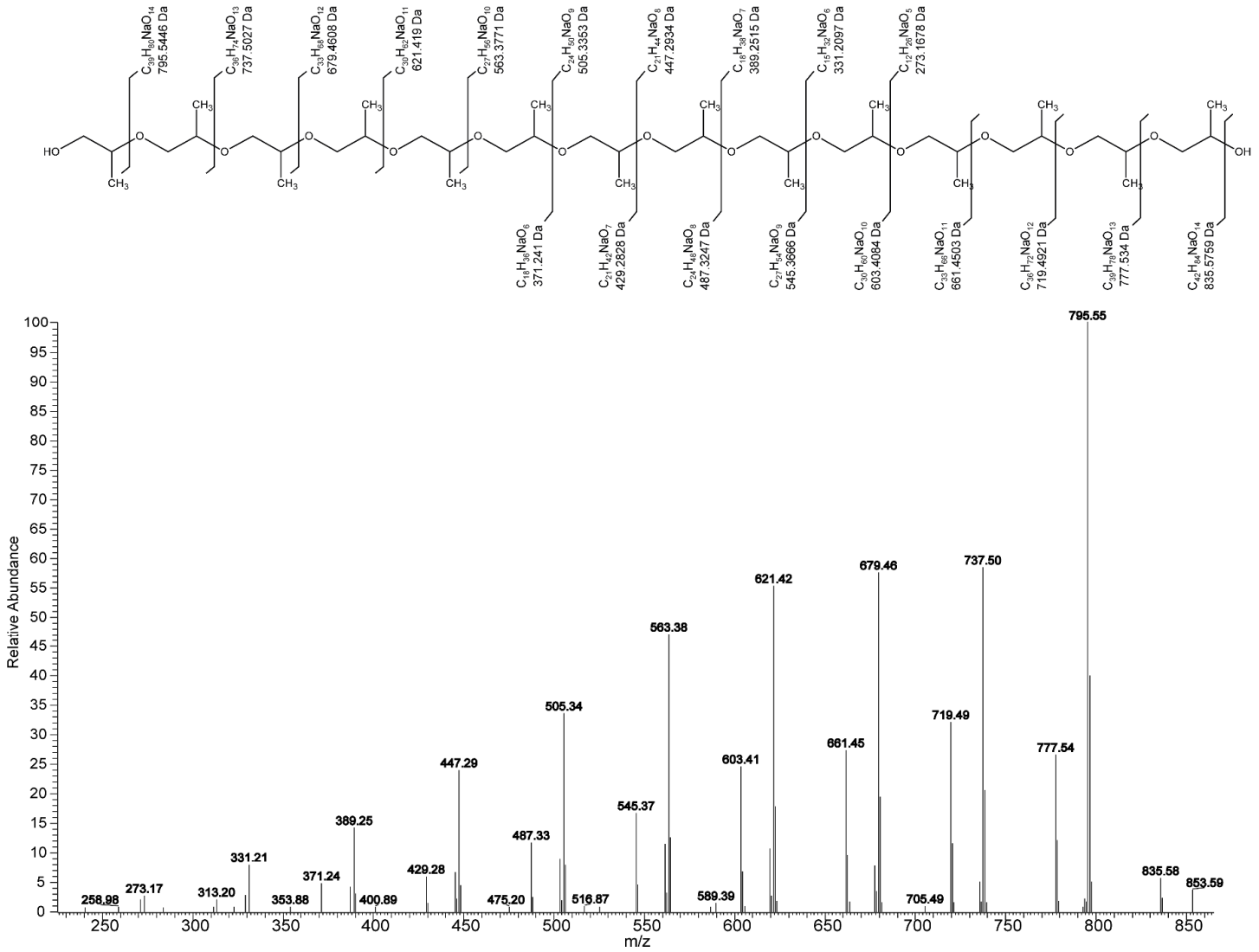


Figure 4.16. Proposed fragmentation pattern of the PPG molecule, associated with series 2 (top), for the recorded product ion spectrum of precursor ion *m/z* 853.59 at retention time 22.91 min CID energy 50 eV.

For the series 4, 5, 10, 11, 12, 14 and 17 and their assigned molecular formulae it was not possible to determine whether any of these series were related by a change of adduct. The EICs of PPG series 11, 12 and 18 showed multiple overlapping peaks and it was not possible to determine a series of peaks that could be identified as PPG. Furthermore, from the recorded product ion spectra of ions from these three series it was not possible to identify any of these series as PPG.

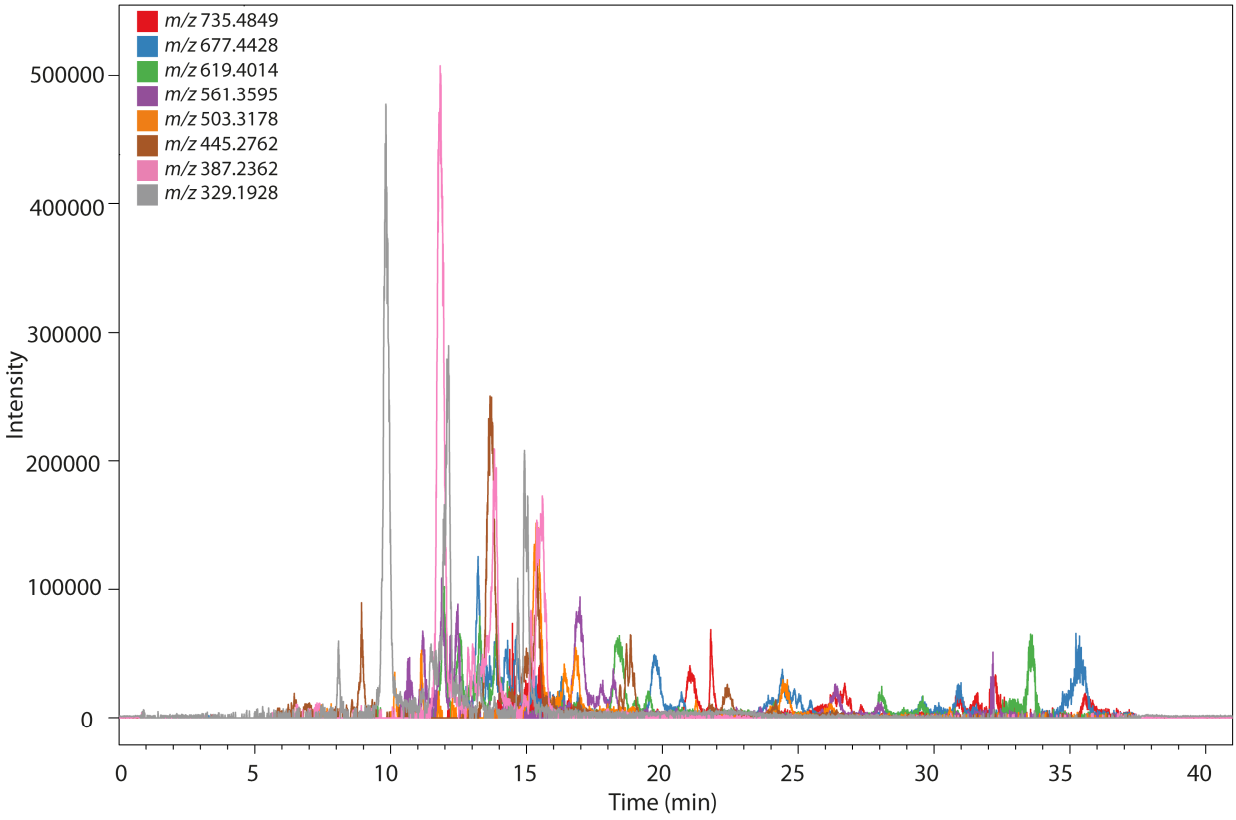


Figure 4.17. EICs from the HPLC-MS analysis of the DOM SPE extract of the sewage effluent. The ions were identified as oligomers of PPG from the DI-HRMS spectra (Table 3.3), the series displayed is PPG series 12.

For the series 4, 5, 10, 14 and 17 the EICs showed clearly resolved peaks which were normally distributed, indicative of a PPG series. For the PPG series 10 and 14 as expected there was a single series of peaks indicative of PPG, the EIC of the oligomers from series 10 are shown in Figure 4.21(a). In contrast, EICs of the oligomeric ions for series 4, 5 and 17 also show resolved normally distributed series of peaks with two peaks associated with each ion (Figure 4.21(b)). This is likely caused by the presence of multiple structural isomers.

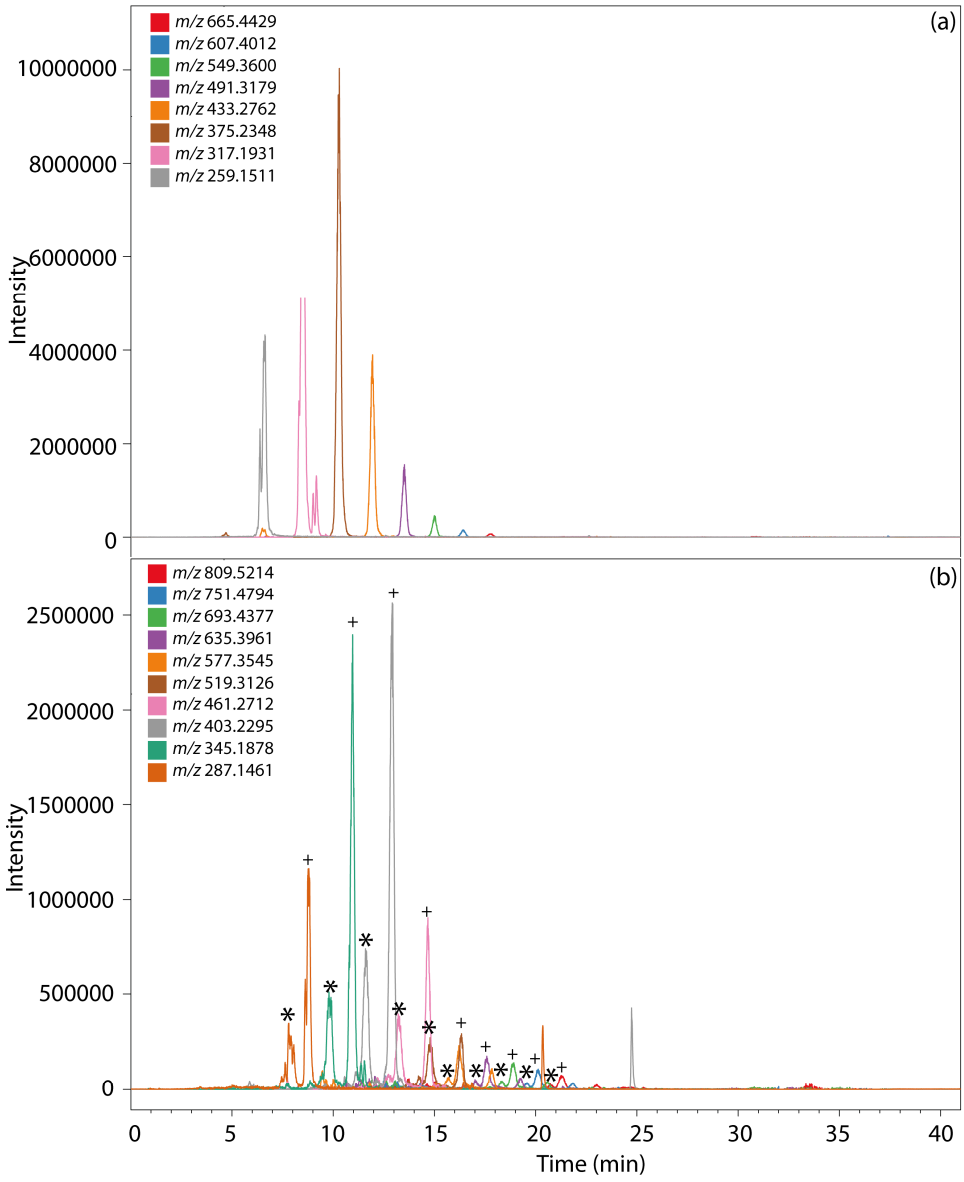
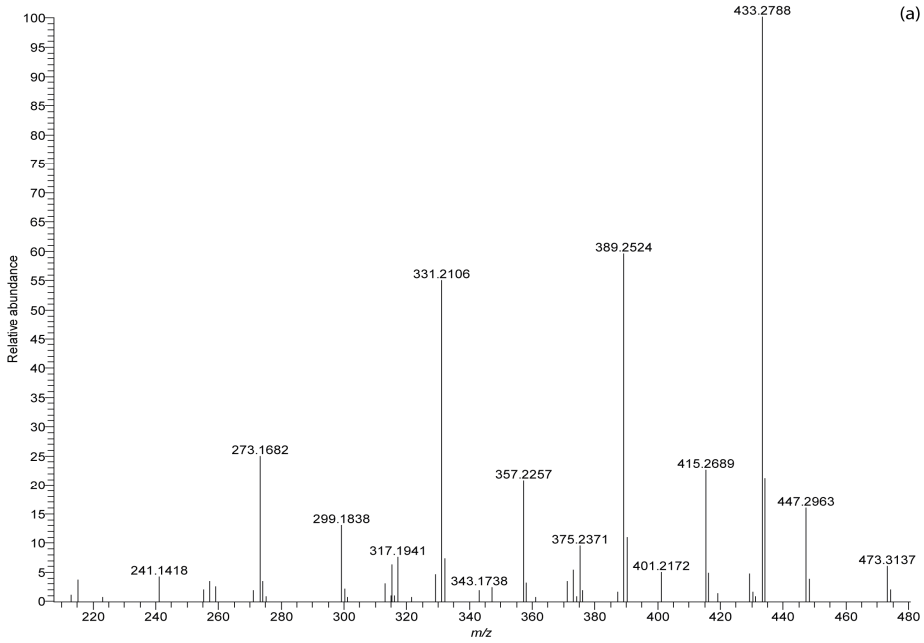
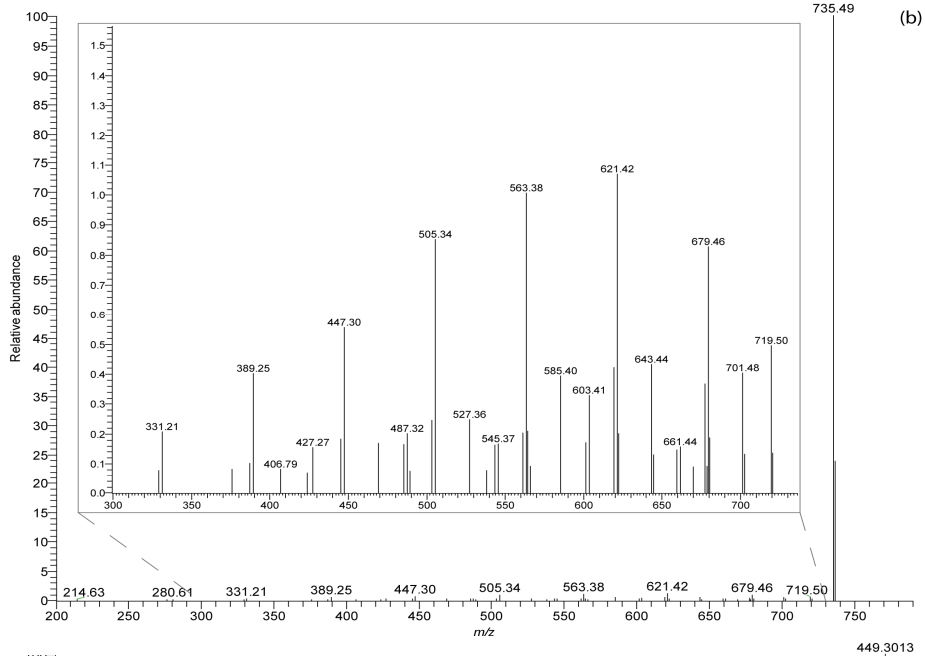
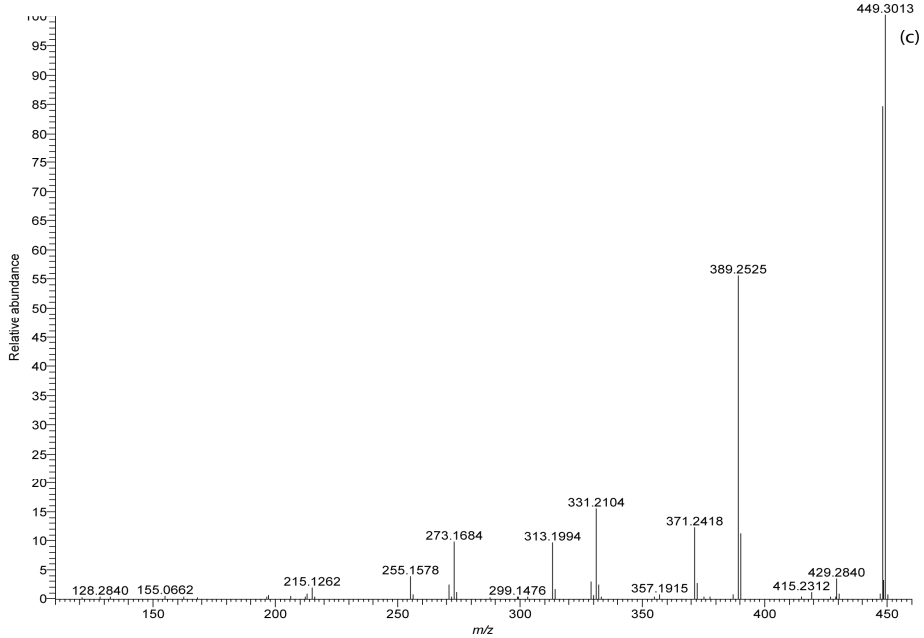


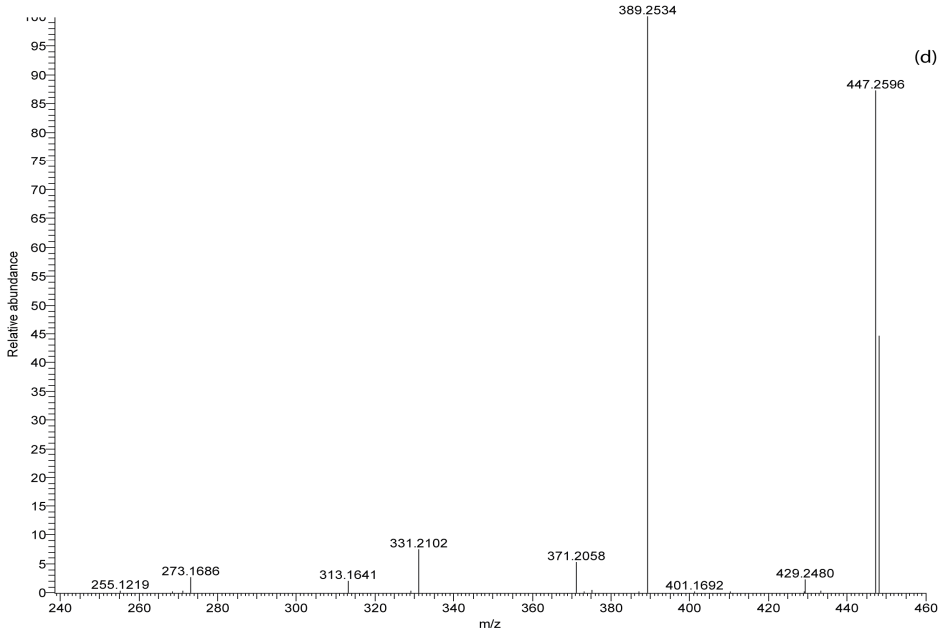
Figure 4.18. EICs from the HPLC-MS analysis of the DOM SPE extract of the sewage effluent. The ions were identified to be oligomers of PPG from the DI-HRMS spectra (Table 3.3), the series displayed are: (a) PPG series 10 (b) PPG series 5 stars highlight the earlier eluting series + highlight the later eluting series.

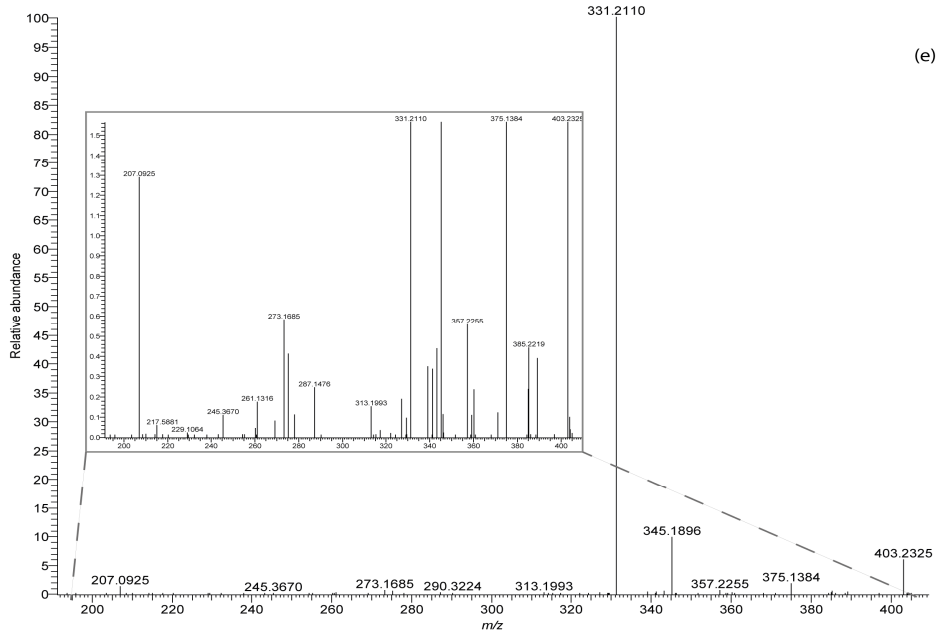
Figure 4.19 shows examples of the product ion spectra of an ion from each of the PPG series, 4, 5, 10, 14 and 17. For the PPG (series 4, 5 and 17 two product ion spectra are shown corresponding to two of the peaks detected). There are clearly series of ions with a 58 Da spacing in each of the spectra shown in Figure 4.19, which shows that all the series have the polymer backbone PPG. However, from the product ion spectra it is not possible to determine the structure of the end groups. This would require a reference standard for each of the series for which both the product ion spectra and the retention time could be used to confirm the structure.

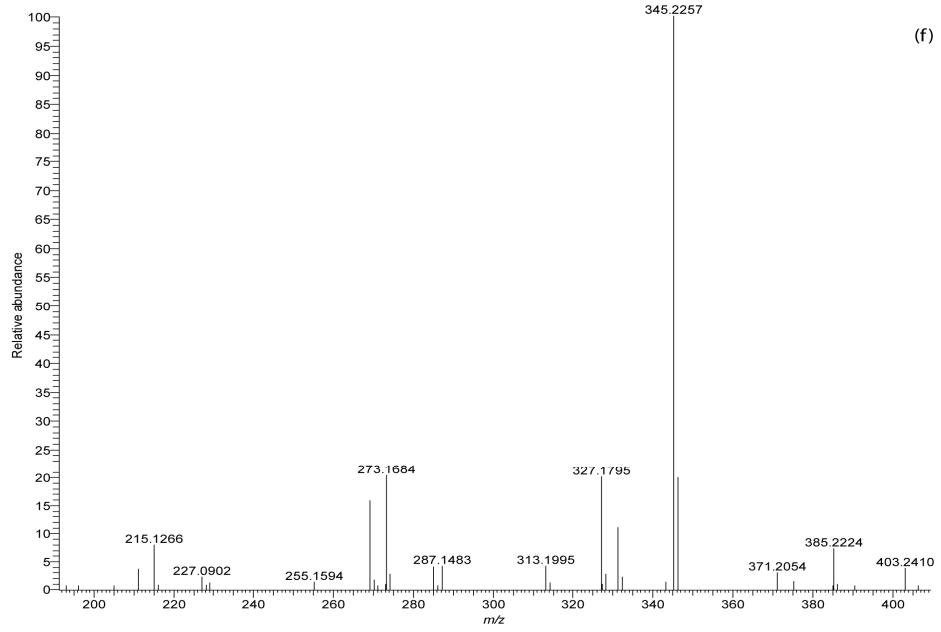

(a)
433.2788
389.2524
331.2106
273.1682
357.2257
415.2689
447.2963
299.1838
375.2371
317.1941
401.2172
473.3137
241.1418
343.1738
Relative abundance
m/z


(b)
735.49
621.42
563.38
505.34
679.46
447.30
719.50
643.44
389.25
585.40
701.48
603.41
527.36
331.21
487.32
545.37
427.27
661.44
406.79
214.63
280.61
331.21
389.25
447.30
505.34
563.38
621.42
679.46
719.50
Relative abundance
m/z


(c)
449.3013
389.2525
331.2104
371.2418
273.1684
313.1994
255.1578
429.2840
215.1262
415.2312
128.2840
155.0662
299.1476
357.1915
Relative abundance
m/z


(d)
389.2534
447.2596
331.2102
371.2058
273.1686
313.1641
255.1219
401.1692
429.2480
Relative abundance
m/z


(e)
331.2110
345.1896
403.2325
207.0925
245.3670
273.1685
290.3224
313.1993
357.2255
375.1384
Relative abundance
m/z


(f)
345.2257
273.1684
327.1795
215.1266
385.2224
287.1483
313.1995
227.0902
255.1594
371.2054
403.2410
Relative abundance
m/z

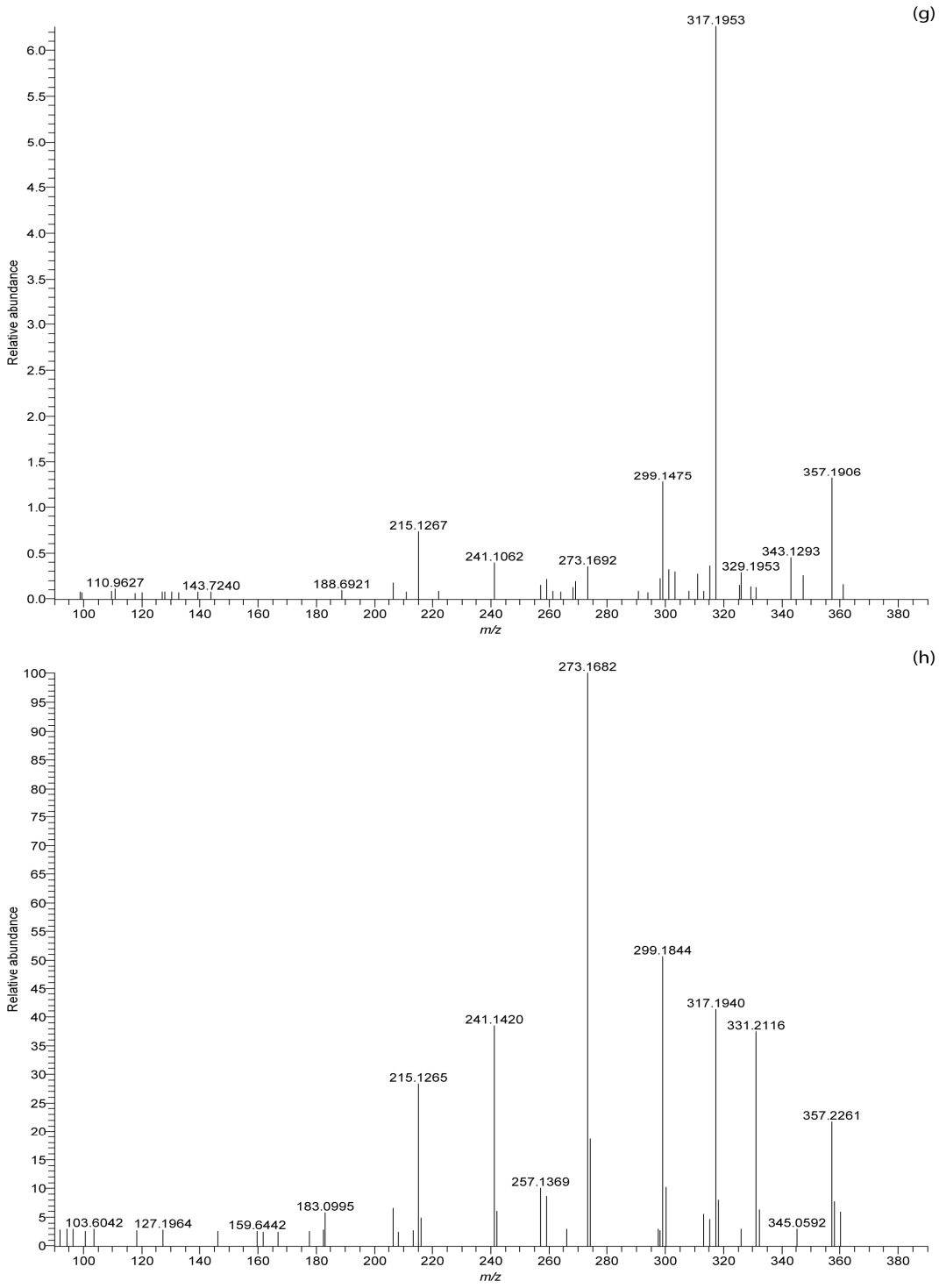


Figure 4.19. (a) Recorded product ion spectrum of precursor ion *m/z* 491.32 from PPG series 10, retention time 13.50 min CID energy 50 eV, (b) recorded product ion spectrum of precursor ion *m/z* 777.531 from PPG series 10, retention time 13.50 min CID energy 50 eV, (c) recorded product ion spectrum of precursor ion *m/z* 477.26 from PPG series 4, retention time 12.19 min CID energy 50 eV, (d) recorded product ion spectrum of precursor ion *m/z* 477.26 from PPG series 4, retention time 13.50 min CID energy 50 eV, (e) recorded product ion spectrum of precursor ion *m/z* 403.23 from PPG series 5, retention time 11.73 min CID energy 50 eV, (f) recorded product ion spectrum of precursor ion *m/z* 403.23 from PPG series 5, retention time 12.96 min CID energy 50 eV, (g) recorded product ion spectrum of precursor ion *m/z* 375.20 from PPG series 17, retention time 9.35 min CID energy 50 eV, and (h) recorded product ion spectrum of precursor ion *m/z* 375.20 from PPG series 17, retention time 10.30 min CID energy 50 eV.

Using the characteristic product ion spectra of PPG and the distribution of oligomers in the EIC, 141 peaks were identified as PPG. Of the 18 series determined from the DI-HRMS 5 series were found to be the PPG diol. The differences in mass were due to different adducts and isotopes of the same compound, identifiable based on coincident peaks in the different ion chromatograms. Three series were identified as the PPG with a hydroxyl and butyl end groups. The difference in retention times between PPG diol and PPG with the hydroxyl and butyl end groups, shows the effect the end group can have on retention of polymeric materials on the C_{18} column. Even though other series showed product ions with a 58 Da mass defect characteristic of PPG it was not possible to conclusively identify the end groups of these oligomeric series. However, using standards of PPG with differing end groups it may be possible to identify these polymers.

The identification of these polymers show the benefit of using DI-HRMS and HPLC-HRMS to analyse SPE extracts. Using the DI-HRMS spectra it was possible to determine series of PPG ions from the spectra with a 58 Da mass defect. These oligomeric series were not apparent in the TIC-HRMS which, due to the high amounts of co-elution. Only by using the exact mass of the ions from series identified from the DI-HRMS was it possible to use the EICs to reveal series of normally distributed peaks indicative of a polymer.

Bacterial cultures have been shown to be able to utilise PPG as an energy source to grow. However, little research has been carried out to characterise the metabolites of PPG. A study by Kawai & Schink, (1987) analysed the metabolism of di-propylene and observed the dehydrogenation of the alcohol end group to an aldehyde as a metabolite before cleavage of the ether bond. This research was further developed by Hu *et al*. (2008) who cultured *Sphingobium* using PPG 700 diol as a nutrient source. It was found that the bacteria dehydrogenate the alcohol end group before cleavage of the ether bond forming PPG oligomeric metabolites with an aldehyde end group. However, the full metabolic pathway is poorly understood for the polymer. The metabolism of PPG is well-characterised in humans. Under normal conditions PPG dehydrogenates to lactaldehyde catalysed by the enzyme alcohol dehydrogenase and then lactic acid by the enzyme aldehyde dehydrogenase and broken down via the Krebs cycle (Niedworok and Rehme, 2014). Because PPG can be microbially degraded, it was surprising to find these compounds persisting through the secondary treatment process of Chew Stoke sewage treatment works. Even though PPG has not been found to be toxic, as a bioavailable compound could be a factor in the cause of eutrophication. Therefore, bioavailable compounds may have an indirect impact on the environment. This is meant to be

mitigated by the biological processing within sewage treatment works. This could be an indication that either the secondary treatment process is not working properly or that higher molecular weight compounds require a longer treatment time to remove them. This shows that even biodegradable compounds persist through the sewage treatment. Furthermore, when bacteria metabolise PPG, oligomeric metabolites with different end groups are formed. This may indicate that some of these PPG series are metabolites of the degradation process. However, confirmation of this, would require further investigation of the microbial degradation of PPG.

4.4 Summary

Using HPLC-HRMS further showed the complexity of the composition of the SPE extracts. Chromatography increased the number of components detected both by (i) decreasing competitive ionisation and increasing sensitivity, and (ii) separating isomers chromatographically (aim 1). The use of the ternary plots to compare the peak areas allowed trends to be determined regarding the origin of the compound (aim 2). Using the exact masses of the ions determined to be statistically significant from the Kruskal-Wallis of the DI-HRMS spectra and their position on the ternary plot shows that these components are predominantly associated with the sewage outfall and downstream extracts (aim 3). The MS^2 experiments allowed the identification of 96 components to be identified (aim 4). Twenty-Two of the compounds had been previously identified either in surface water or sewage effluent showing that the Kruskal-Wallis analysis highlights compounds expected to originate from a sewage outfall (aim 5, hypothesis 1). Two compounds were identified not previously found in sewage effluent showing the benefit of using an untargeted methodology (hypothesis 2).

4.5 Conclusions

This chapter presents the results of analysing the composition of the SPE extracts by HPLC-ESI-Orbitrap-HRMS and HPLC-ESI-Orbitrap-HRMS/MS. HPLC-MS has further shown the complexity of the SPE extracts to investigate the aims and hypotheses stated. The major findings are that:

(i) Due to the complexity and chemodiversity of the compounds in the DOM extracts the TIC shows extensive co-elution. However, EIC of individual ions showed that compounds are separated chromatographically.

(ii) The peak picking highlighted the complexity of the SPE extracts finding over 14325 EIC peaks across the SPE sample extracts. (aim i)

(iii) Using ternary plots comparing the ratio of the peak areas for individual components between the SPE extracts. Trends could then be observed including which components were found in the sewage outfall and downstream extracts and not in the upstream, and the components unique to the riverine extracts. (aim ii)

(iv) It was found that of the 510 ions determined to be statistically significant and coming from the sewage outfall, 420 were detected in the HPLC-MS analysis. The 420 ions detected corresponded to 681 chromatographic peaks. (aim iii)

(v) Using the ternary plot, it was shown that the majority plotted in the area corresponding to the ions found predominantly in the sewage outfall and downstream extracts. This shows that using the DI-HRMS spectra and the Kruskal-Wallis analysis the statistically significant ions were found to predominantly be coming from the sewage outfall. (aim iii)

(vi) Twenty-two compounds were identified originating from the sewage outfall including pharmaceuticals, plasticisers, metabolites and illicit drugs. Many have been identified in previous studies as originating from sewage treatment works. This shows that this method highlights a relevant list of compounds coming from a source without the need to determine a target list of compounds prior to analysis. (aim iv & v)

(vii) Two compounds were found which have not been identified as originating from sewage outfalls in previous studies, namely raltegravir and piperine. Raltegravir is an antiretroviral, which is used in the treatment of HIV and AIDS. Due to the low prevalence of HIV and AIDs in the UK it was surprising to find this in a sewage treatment works. This shows the limitation of targeting or screening for a specific list of compounds. Piperine is the natural product of black pepper which is not native to the UK but commonly used in cooking. (aim iv)

(viii) Using the EICs from the HPLC-MS of the sewage outfall extracts, ions tentatively characterised as PPG from the DI-HRMS spectra could be identified.

(ix) Upon fragmentation 141 of the EIC peaks attributed to PPG were found to have series of 58 Da spaced product ions. This confirms that the polymeric backbone of these compounds is PPG.

(x) Two of the PPG series end groups were able to be identified; the diol and, the hydroxyl and butyl end groups. The different end groups of the PPG series maybe

a result of the industrial synthesis as discussed in Chapter 3 or may be metabolites formed in the processing of the sewage.

Overall, the combination of using both DI-HESI-Orbitrap-HRMS and HPLC-HESI-Orbitrap-HRMS allows a complementary analysis of these complex DOM extracts to determine which compounds were originating from the sewage outfall. This study showed that the ions identified as statistically significant in the DI-HRMS datasets can be identified. Some of the compounds had been previously identified using targeted analysis as originating from sewage outfalls. This shows that this prioritisation method picks compounds relevant to the point source being analysed. In addition, this method identified two compounds not previously characterised from sewage treatment works raltegravir and piperine. Furthermore, using the DI-HRMS analysis allowed the identification of series of PPG ions, which would not have been possible to determine using only the HPLC-MS analysis.

However, with many sewage treatment works in the UK alone which differ in environmental setting, treatment method and population demographic. Therefore, this method needs to be further tested in different environmental settings to determine its applicability to other point sources. Such examples are presented in the following chapters.

Chapter 5

Untargeted analysis of compounds originating from Betws-y-Coed sewage treatment works

Chapter 5. Untargeted analysis of compounds originating from Betws-y-Coed sewage treatment works

5.1 Introduction

There are over 9,000 sewage treatment works in the UK alone (DEFRA, 2012). As outlined in Chapter 1, different sewage treatment works use different treatment methods which may affect the chemical composition of the resulting treated effluent. Furthermore, the sewage entering the treatment works will depend on the population demographic, the industries, and the size of the population discharging waste into the sewage network. This in turn will affect the composition of the treated sewage being discharged into the environment (DEFRA, 2002, 2012).

As the water travels through the landscape it transports organic compounds from soils and drainage systems into streams and rivers (e.g. Aitkenhead *et al.*, 1999; Stedmon, Markager and Bro, 2003; Baker and Spencer, 2004; Lloyd *et al.*, 2012; Dubinenkov *et al.*, 2015; Roth *et al.*, 2015). Hence, the DOM entering river and lake systems varies depending on the landscape character including topography, soil type, land use and a wide range of anthropogenic sources, including agrichemicals, run-off from roads and storm drains (e.g. Chen *et al.*, 2005; Arias-Estévez *et al.*, 2008; Chadwick, 2011; Chopra *et al.*, 2011; Kaczala and Blum, 2016; Baker, 2004; Stedmon and Bro, 2008; Dubinenkov, 2015). Further contributions to riverine DOM will include multiple point sources such as sewage treatment works, septic tanks, and industrial sites (e.g. Lefebvre and Moletta, 2006; Ying, 2006; Zhang, Hibberd and Zhou, 2008; Goel and Kaur, 2012; Loos *et al.*, 2012; Dvořák *et al.*, 2014; Ruff *et al.*, 2015). These either directly discharge into the river or are washed into a river by precipitation and irrigation (Chen *et al.*, 2005; Arias-Estévez *et al.*, 2008; Lloyd *et al.*, 2012). Therefore, compounds derived from these sources will increase the chemodiversity of the DOM in the river. These various contributions will make determining which compounds originate uniquely from a given point source in a catchment difficult to determine, particularly, if compounds from a given point source may already be present in the receiving river DOM. This may also affect if it is possible to determine whether the source has changed the DOM composition downstream of the point source.

As discussed in Chapter 3 Chew Stoke sewage works was selected as a test site to develop a new approach to determining the composition of DOM contributed to rivers by sewage works based on DI-HRMS, HPLC-Orbitrap-MS and statistical data processing methods. Chew Stoke

was chosen as the River Chew is of modest size and the DOM derives solely from Chew valley lake. This latter feature had advantages as lake water has been shown to display lower chemodiversity than other aquatic environments (Kellerman *et al*., 2015) thereby simplifying the subsequent processes of deconvoluting riverine and sewage work DOM contributions. Having established the new methodologies in the previous chapters the next step in the application of the approach was to begin to test its utility to sewage works located on other river systems of different flows and catchment characters.

As has already been shown in previous targeted studies of different sewage treatment works the presence or absence and concentration of compounds in sewage effluent and surface water varies enormously (Loos *et al.*, 2012; Ruff *et al.*, 2015). However, untargeted analyses of DOM may provide a different picture of the chemical character of DOM in terms of qualitative and quantitative differences and/or similarities in overall compositions. The next two chapters will explore this question in detail.

5.2 Aims

The overall aim of this chapter is to determine whether the methods developed in Chapters 3 and 4 applied to the Betws-y-Coed sewage treatment works in North Wales situated in the Conwy catchment will provide chemical signatures reflecting catchment character and sewage effluent contributions.

The specific aims of this chapter are to:

(vi) Explore the differences in the high-resolution DI-HRMS spectra of the riverine water DOM from the Afon Llugwy and the sewage effluent from Betws-y-Coed sewage treatment plant.

(vii) Determine if the composition of the DI-HRMS spectra of the SPE extracts are significantly different.

(viii) Use the Kruskal-Wallis analysis to determine which ions are significant and discriminating between the composition of the upstream and the sewage effluent SPE extracts.

(ix) Determine if there are any trends in the HPLC-MS data using a ternary plot comparing the peak areas of EICs across different SPE extracts.

(x) Identify compounds found to be statistically significant based on (iii) above.

It is hypothesised that: (i) the molecular "fingerprint" of the extracted sewage effluent from Betws-y-Coed sewage treatment works will be similar to the SPE extract of sewage effluent from Chew Stoke. This is hypothesised as similar products are likely to be used in UK households which will be discharged as sewage. Therefore these compounds are likely to be found at most sewage treatment works. (ii) the molecular "fingerprint" of the upstream DOM SPE extract from the Afon Llugwy will show a similar complex mixture of compounds to the river Chew as previous literature shows a complex mixture of ions from other riverine sources (Osborne et al., 2013; Minor et al., 2014)

5.3 Site information

Betws-y-Coed is situated on the confluence of the Afon Llugwy and the River Conwy and sewage effluent is discharged into the Afon Llugwy (Figure 5.1). The sewage works uses active sludge as its secondary treatment and no tertiary treatment method is used.

The Afon Llugwy catchment upstream of the Betws-y-Coed sewage treatment works encompasses an area of land approximately 110 km^2 highlighted in red in Figure 5.1. The Afon Llugwy drains from the upland peat landscape and runs through the Snowdonia national park. However, this river represents a small part of the overall Conwy catchment. There are several small villages situated along the banks of the Afon Llugwy. The domestic sewage from these villages are either treated using small package sewage treatment works or septic tanks. The largest village upstream of Betws-y-Coed is Capel Curig which has a small treatment works which discharges into the Afon Llugwy. Downstream of Capel Curig is an army cadet training camp which also has its own sewage treatment works. The sub-catchment of the Afon Llugwy contains only low levels of industry and agriculture upstream of the sewage treatment works because the area is protected as part of the Snowdonia National Park. The industries upstream of Betws-y-Coed include a small number of timber mills and hotels, and low levels of mixed livestock farming. All these sources have the potential to contribute to the composition of the DOM in the Afon Llugwy upstream of Betws-y-Coed.

Water samples were collected as described in Chapter 2 directly from the effluent discharge from the Betws-y-Coed sewage treatment works, approximately 140 m upstream and downstream.

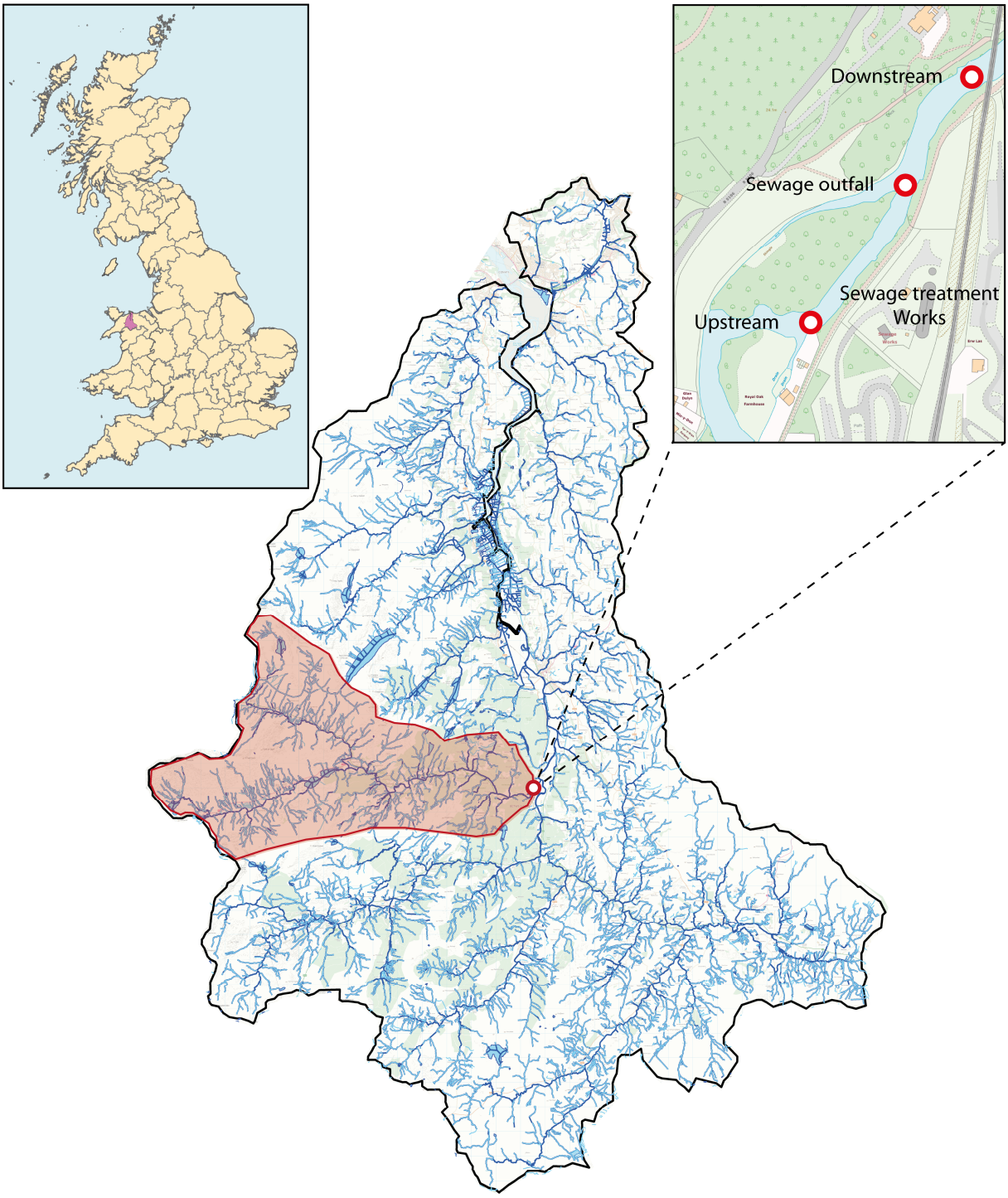


Figure 5.1. Map of the UK with the Conwy catchment area highlighted in pink. Map of the Conwy catchment area and river network. Red highlighted the Afon Llugwy sub-catchment upstream of the Betws-y-Coed sewage treatment works (O). A local map of the sampling sites at Betws-y-Coed sewage treatment works.

5.4 Analysis of the composition of DOM extracts in Betws-y-Coed

The water samples and comparative HPLC water were filtered and extracted as described in Section 2.2, within 48 h of collection. The dry extracts were dissolved in methanol/water (1:1

v/v, 4.5 ml). There was no visual difference between the riverine water samples. However, the sewage effluent was noticeably coloured and had a strong odour. The DOM extracts of the water collected upstream and downstream were yellow coloured and clear. The extracts from the sewage outfall was noticeably darker than the riverine DOM extracts. The blank extracts were clear and colourless.

The volume of water extracted, the extraction method and the volume of solvent used to dissolve the SPE extracts was consistent for all samples in order allow direct comparison of the composition of the DOM in each extract.

5.4.1 DOC analysis of water samples and SPE extracts from Betws-y-Coed

DOC analysis was carried out as described in Section 2.3.1 on filtered water from the upstream, downstream and sewage outfall. The concentration of organic carbon in the filtered water collected from each site, the concentration of organic carbon in the SPE extracts and the extraction efficiency are displayed in Table 5.1. The results show that the concentration of dissolved organic carbon in the filtered riverine water samples is three times lower than that found in the sewage outfall. However, the two riverine concentrations are similar, showing that the sewage outfall is not changing the overall concentration of organic carbon in the river.

Table 5.1. DOC of the water sampled from each site, The amount of organic carbon in each SPE extract Extraction efficiencies of the SPE extraction.

Samples	**DOC filtered water (mgl^{-1})**	**DOC extracts (mg)**	**Extract efficiencies (%)**
Upstream	2.11 ± 0.03	1.69 ± 0.10	80.07 ± 5.0
Sewage Outfall	7.12 ± 0.03	3.81 ± 0.31	53.47 ± 3.7
Downstream	2.21 ± 0.02	1.75 ± 0.08	79.23 ± 4.3
Blank	N/A	0.28 ± 0.02	N/A

The extraction efficiencies were higher in the riverine extracts than the sewage outfall wastewater extracts. This is likely because the composition of the wastewater and riverine DOM differ and it is well-known from the literature that individual compounds can have different extraction efficiencies using the same extraction method (Raeke *et al.*, 2016; Li *et al.*, 2017; Arellano *et al.*, 2018).

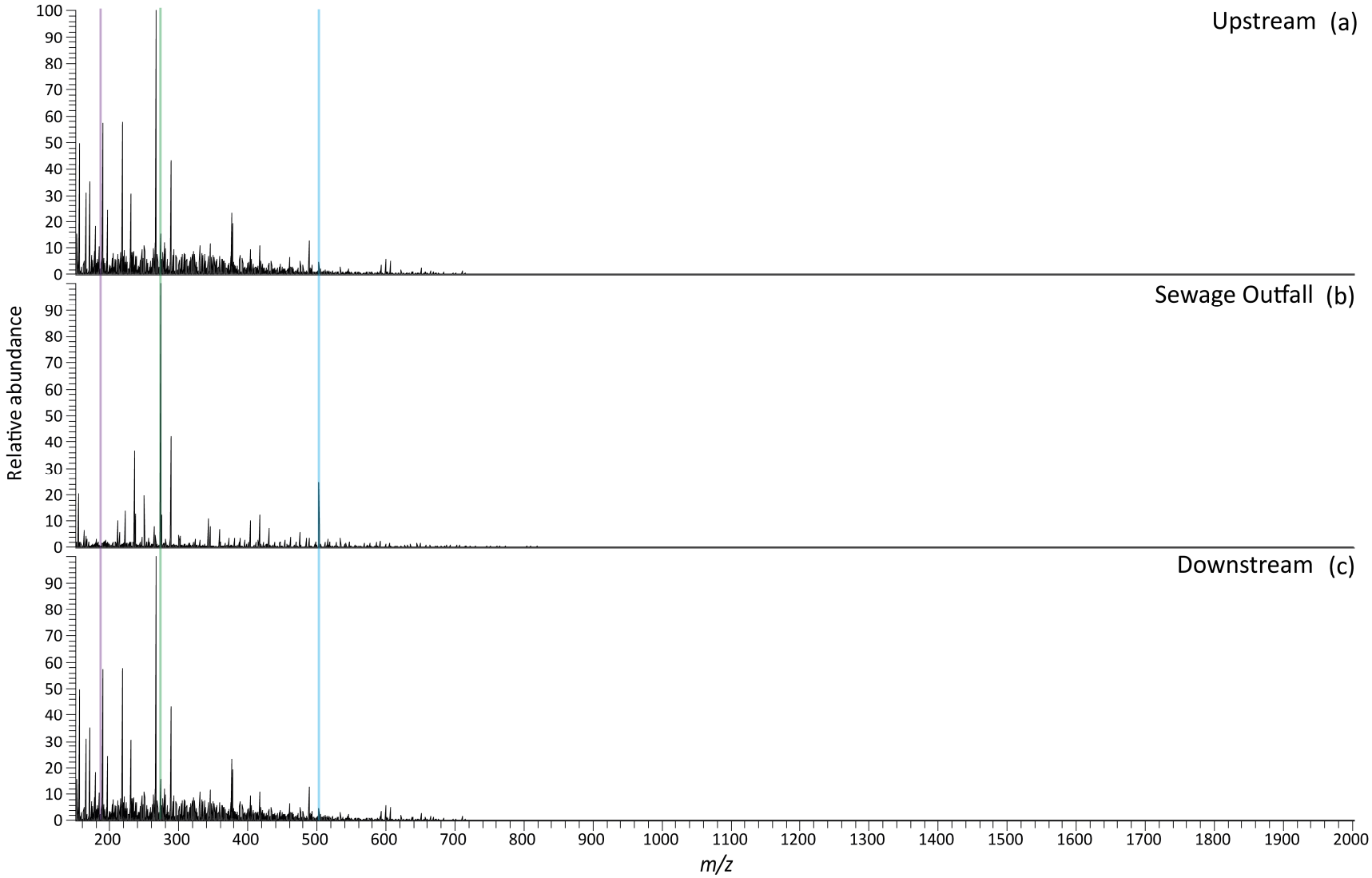


Figure 5.2. DI-HRMS spectra of DOM SPE extracted from the: (a) upstream, (b) sewage outfall and, (c) downstream water samples collected from Betws-y-Coed. The mass ranges highlighted correspond to the mass ranges used in Figure 5.4; purple *m/z* 206.8 to 209.2, green range *m/z* 273.0 to 276.4 and blue range *m/z* 501.0 to 504.5.

5.4.2 DI-HRMS analysis of SPE DOM extracts from Betws-Y-Coed

All of the replicate SPE extracts collected from the Betws-y-Coed sampling sites, and the blank were analysed using DI-HRMS methods outlined in Section 2.3.2 shows a DI-HRMS spectrum of the DOM extracted from the upstream, downstream and sewage outfall from Betws-y-Coed for the range *m/z* 150 to 2000. The spectra show that there is are clear compositional differences between the DOM extracted from the river and the DOM extracted from the sewage outfall. The composition of the downstream DOM SPE extract does not appear to have been affected by the sewage outfall, with the upstream and downstream DI-HRMS spectrum appearing very similar, with the later showing no obvious ions common to the sewage outfall DI-HRMS spectrum.

The upstream DOM at Chew Stoke showed a characteristic complex DI-HRMS mass spectrum characterised by clusters of isobaric ions at 1 Da intervals with an overall normal distribution of intensities typically seen in lake and river DOM extracts (Osborne et al. 2013; Mangal et al. 2016; Flerus et al. 2011). The Betws-y-Coed upstream DOM DI-HRMS spectrum shown in Figure 5.3 reveals a similar complex pattern of ions as the DI-HRMS spectrum of the upstream DOM from Chew Stoke (Figure 3.2).

In contrast, comparing the sewage outfall DI-HRMS spectra from Betws-y-Coed and Chew Stoke, showed the former to lack the clear series of ions indicative of the homologous series of compounds which were such a prominent feature of the DI-HRMS spectra at Chew Stoke (Figure 3.7). Betws-y-Coed sewage outfall's composition instead shows several discrete ions which are more abundant than other ions in the spectrum. This suggests that the composition of the DOM from sewage outfalls may vary significantly between different works rather than

displaying a characteristic point source “fingerprint” disproving the hypothesis stated above.

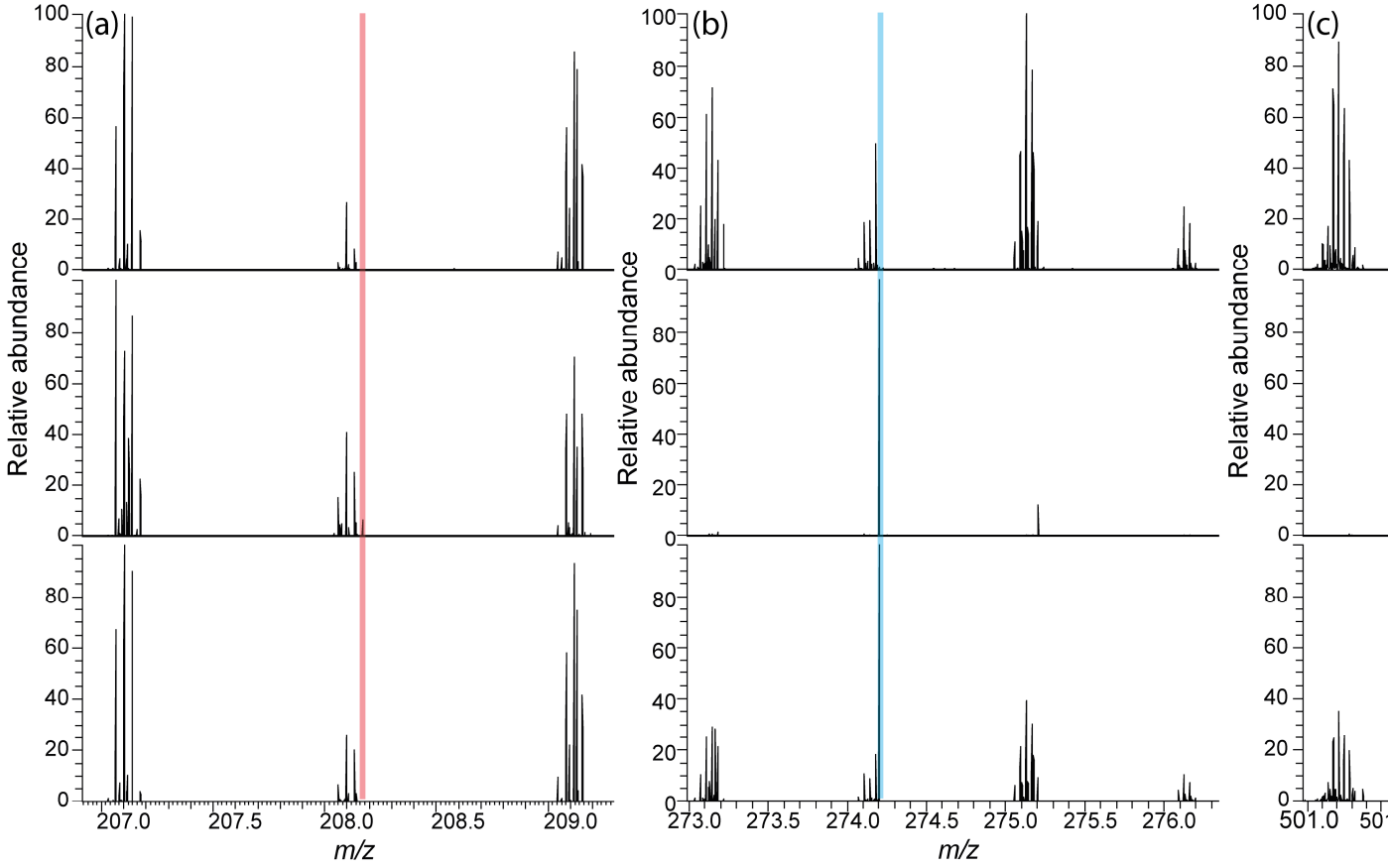


Figure 5.4 shows a comparison of the DI-HRMS spectra of the upstream, downstream and sewage outfall SPE extracts displaying three narrower mass ranges to allow comparison of individual ions. Figure 5.4(a) shows a 3 Da mass range between *m/z* 206.8 to 209.2. Despite the DI-HRMS spectra of the sewage outfall and the riverine SPE extracts appearing different when comparing the full range of the DI-HRMS spectra, when comparing a narrower mass range shows that similarities do exist in the composition of the three DOM extracts. However, Figure 5.4(a) also shows that when comparing individual ions there is an additional ion in the sewage outfall DI-HRMS spectrum, *m/z* 208.1700 (highlighted in red), not seen in spectra of

the upstream and downstream DI-HRMS spectra.

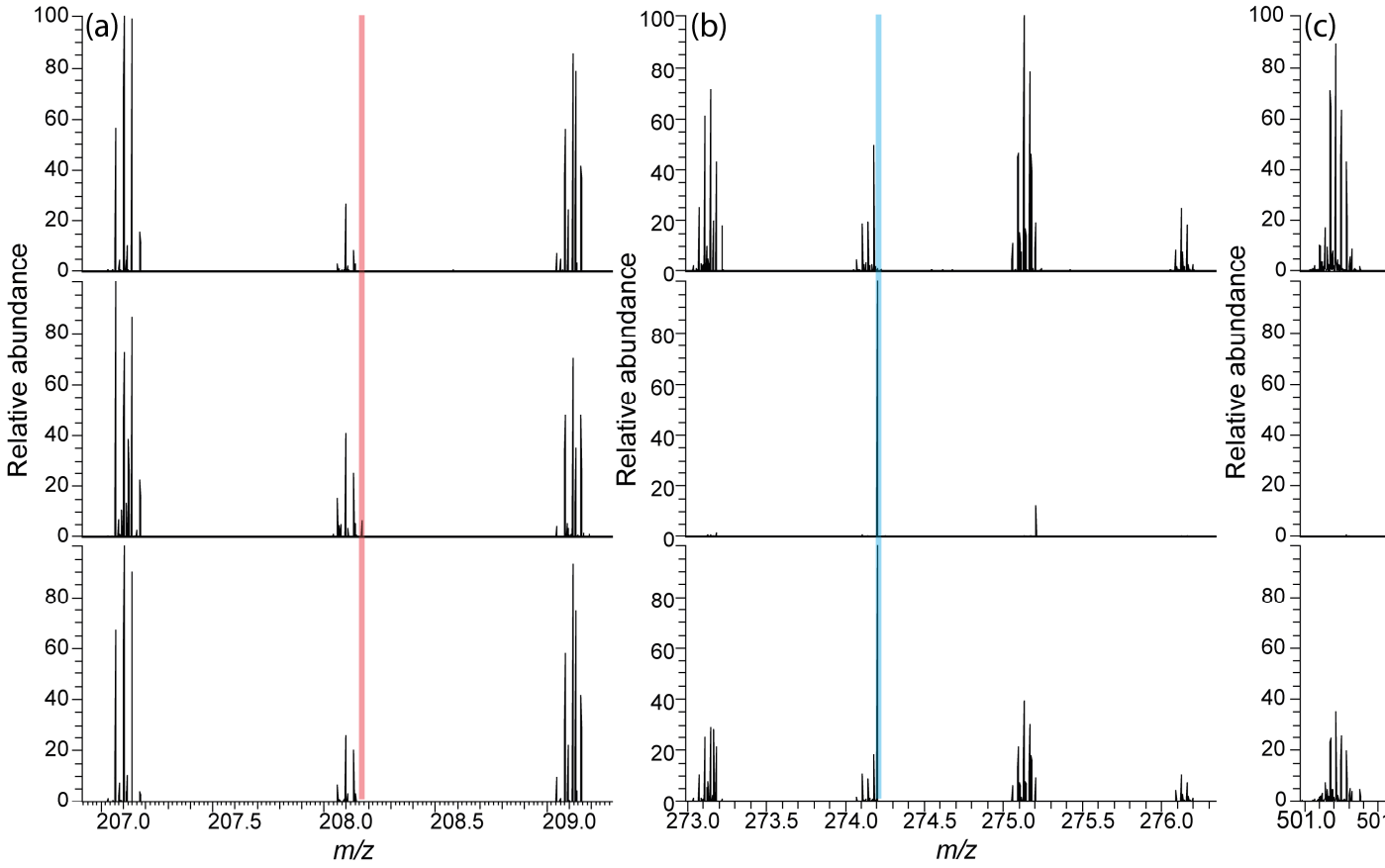


Figure 5.4 (b) and (c) display narrower mass ranges containing two of the most abundant ions in the sewage outfall DI-HRMS spectrum, *m/z* 274.2018 and *m/z* 502.2962 (highlighted in blue and purple respectively). The ions are clearly present in the downstream and sewage outfall DI-HRMS spectra and absent from the upstream DI-HRMS spectra. This shows that the sewage outfall has affected the composition of the downstream DOM despite this not being immediately evident from the full mass range DI-HRMS spectra of the different SPE extracts. The downstream extract clearly also contains ions present in the upstream DI-HRMS spectrum showing, as expected, the DOM composition downstream is a mixture of the riverine and sewage outfall DOM.

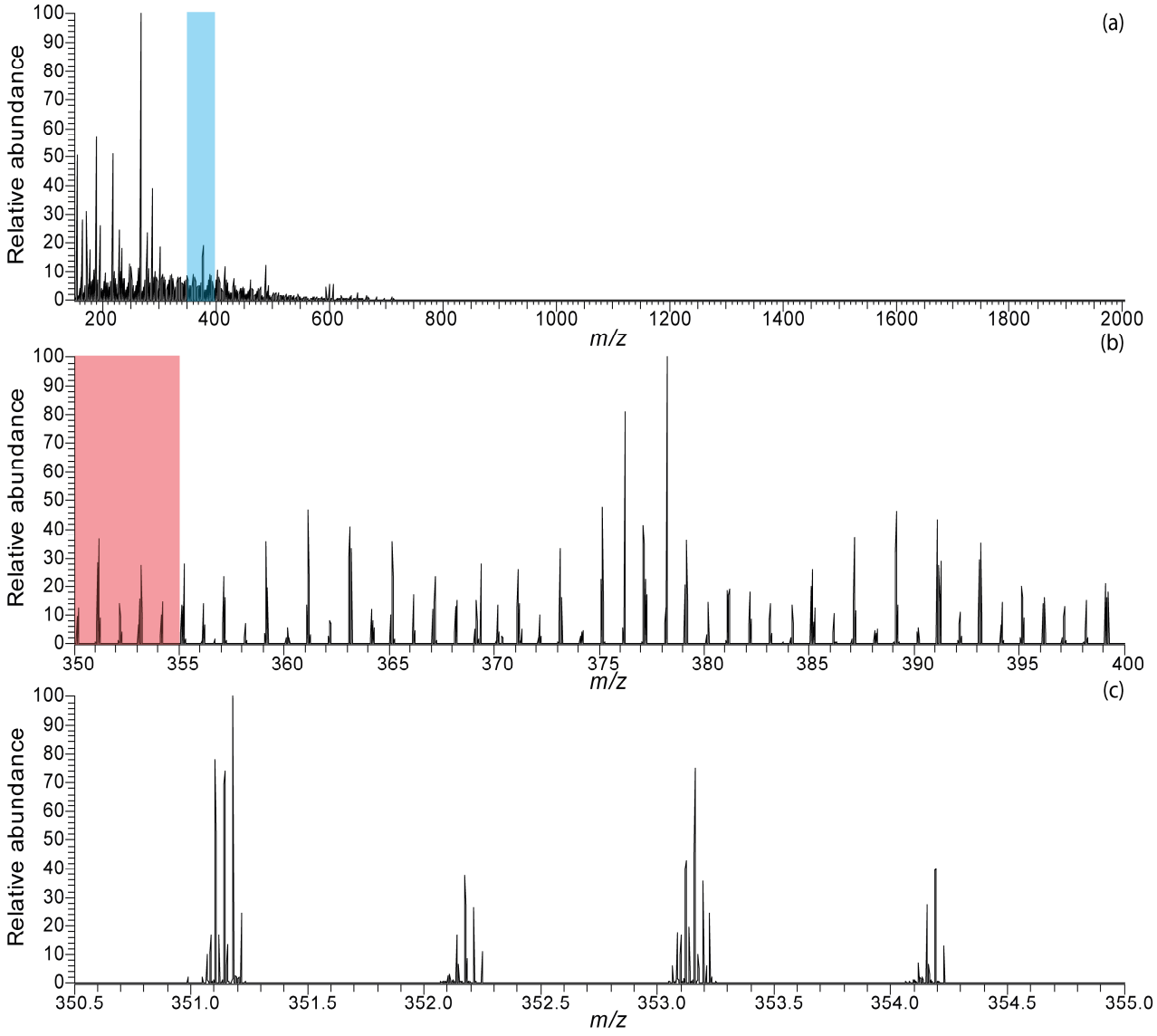


Figure 5.3. DI-HRMS spectra of the upstream DOM extract at displaying three different mass ranges extracted from Betws-y-Coed. DI-HRMS spectra of the upstream SPE extract: (a) full range *m/z* 150 to 2000, highlighted in blue is the 50 Da mass range (b) 50 Da range *m/z* 350 to 400 highlighted in red is the 5 Da mass range, and (c) 5 Da range *m/z* 350.5 to 355.0.

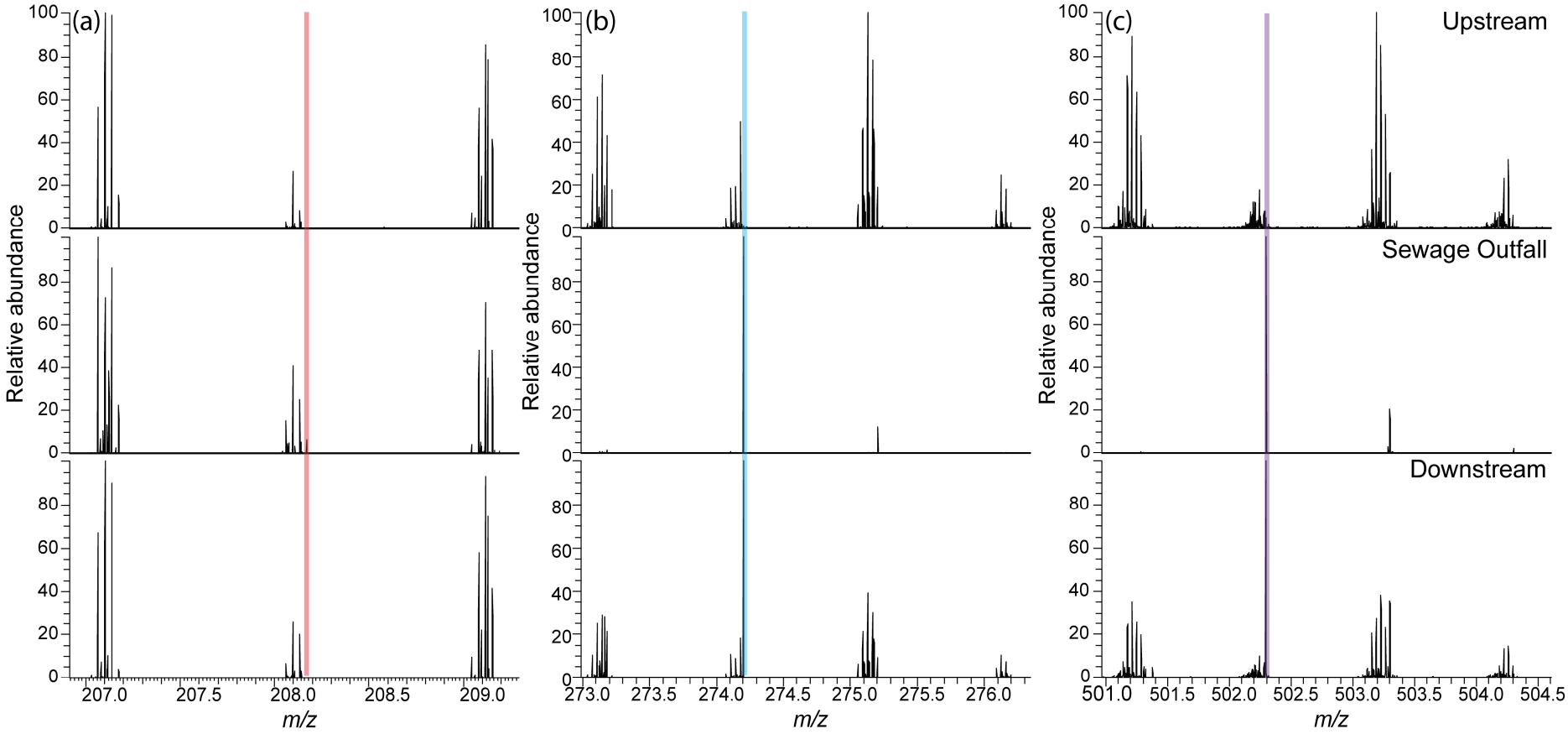


Figure 5.4. Comparison selected narrow mass ranges of DI-HRMS spectra of the upstream, downstream and sewage outfall SPE extracts from Betws-y-Coed: (a) range *m/z* 206.8 to 209.2 the ion *m/z* 208.1700 is highlighted in red, (b) range *m/z* 273.0 to 276.4, the ion *m/z* 274.2018 is highlighted in blue and, (c) range *m/z* 501.0 to 504.5, the ion *m/z* 502.2962 is highlighted in purple.

In summary, the water samples extracted at Betws-y-Coed show a clear compositional difference in the sewage outfall and riverine DI-HRMS spectra. When comparing the full mass ranges of DI-HRMS spectra of the composition of the SPE extracts the sewage outfall does not appear to have affected the downstream DOM composition. However, by analysing narrower mass ranges across all the spectra, there are clearly ions present only in the downstream and sewage outfall spectra, which are not present in the upstream spectra. Furthermore, when comparing individual ions across the sewage outfall, upstream and downstream DI-HRMS spectra there are also ions which are common to all three extracts showing that parts of the composition of DOM SPE extracts are similar. As shown in Chapter 3 it is impractical to comprehensively compare DI-HRMS spectra manually, making the use of chemometric methods essential to determine significant differences between sources.

5.4.2.1 Ion picking and alignment

Ion picking and alignment of all analyses of the downstream, upstream, sewage outfall, blank and QCs extracts were carried out as described in Section 2.3.2.1. The scan to scan variability and variability in the mass accuracy over all analyses was 3.4 ppm and 1.1 ppm respectively. Therefore, the ions were aligned by their exact mass within a 5 ppm range. Across all DI-HRMS spectra 1336 ions were found from the DI-HRMS spectra of the SPE extracts of water collected from Betws-y-Coed above a signal-to-noise ratio of 5.

5.4.2.2 Statistical analysis for Betws-y-Coed

The ion picking, and alignment created a data matrix of ion intensities aligned by mass as described in Section 2.3.2.1 for the Betws-y-Coed SPE extracts. Three statistical analyses were then applied to the DI-HRMS data, PCA, hierarchical, cluster analysis combined with a heatmap of ion intensities, and Kruskal-Wallis one-way analysis.

The PCA displays the differences in the composition of the QC, upstream, downstream, sewage outfall and blank replicate SPE extracts (Figure 5.5). All the extraction replicates cluster in their respective sample groups showing that the extraction and analytical variance is lower than the differences in the composition between the DOM extracts. There is a clear separation in PC1 between the downstream, upstream, sewage outfall and QCs in comparison to the blank HPLC grade water extracts. This represents 54.56 % of the overall variance. PC2 shows a clear separation between the composition of the upstream, downstream, QC and sewage outfall SPE extracts which represents 17.76 % of the variance. The pooled QCs are clustered between the

sewage outfall and downstream as expected. When analysing the Chew Stoke extracts in Chapter 3, the results showed that the mixed QC contained additional compounds input by the sewage outfall which were not found in the upstream extracts. The clustering of the QC indicates that the analytical variance and pre-processing are not influencing the separation of the samples. The PCA shows that the sewage outfall and upstream are least similar in composition. The downstream SPE extracts plot in between the sewage outfall and upstream extracts reflecting the in-stream mixing. These results are similar to those obtained from Chew Stoke. Except that the outfall results in a greater variance between the sewage outfall and riverine at Betws-y-Coed than Chew Stoke. In addition, the downstream extracts plot closer to the upstream extracts which when initially comparing the DI-HRMS spectra the downstream appeared more similar to the upstream spectra.

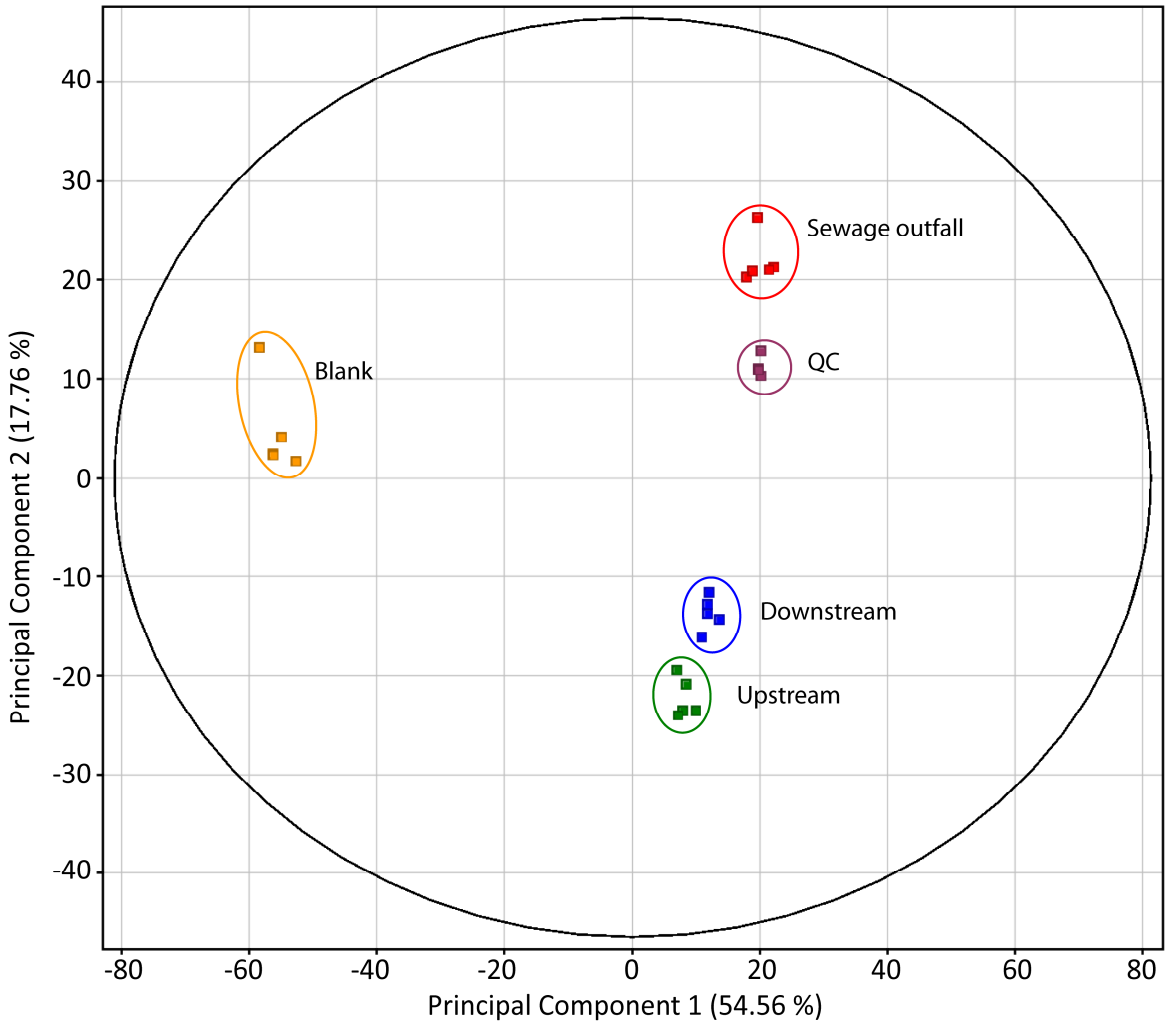


Figure 5.5. PCA analysis of the ions found in the Betws-y-Coed DI-HRMS spectra of the upstream (■), downstream (■), sewage outfall (■), QC (■) and blank (■) SPE extracts.

Figure 5.6 shows the PCA of the upstream, downstream and sewage outfall extracts which were calculated to determine whether the inclusion of the blanks and QC were influencing the separation observed between the SPE extracts of the sampling sites. The PCA shows that the upstream, downstream and sewage outfall separated in PC1 which explains 44.21 % of the variance between these DI-HRMS spectra. This shows that the composition of the sewage outfall and the upstream extracts are least similar. The downstream composition plots between the sewage outfall and upstream. This fits the hypothesis that the downstream DOM is detected by DI-HRMS analysis as an admixture of the upstream and sewage outfall DOMs. This was similar to the findings at Chew Stoke sewage treatment works.

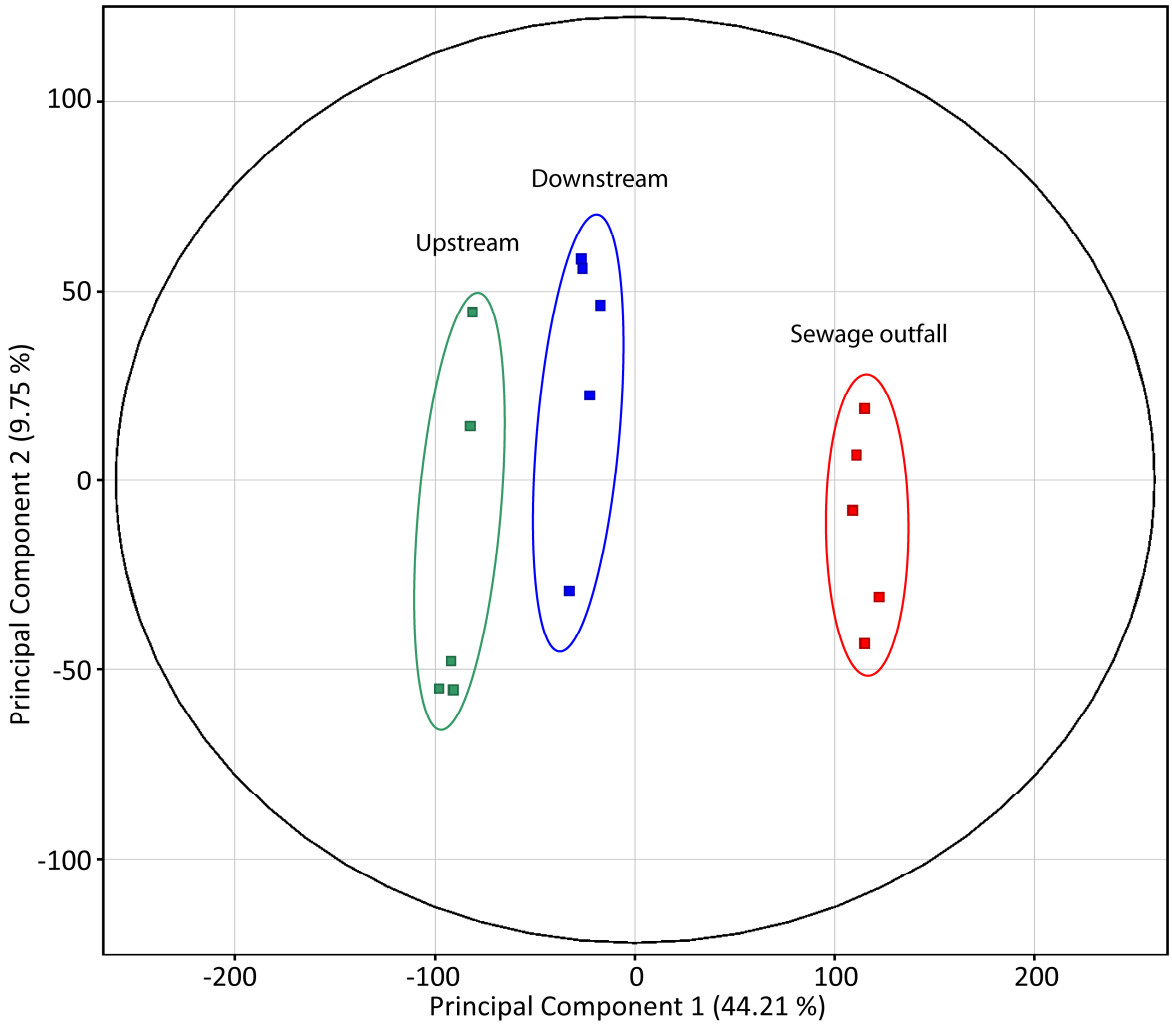


Figure 5.6. PCA analysis of Betws-y-Coed of the DI-HRMS spectra of the DOM extracts from the upstream (■), downstream (■) and sewage outfall (■).

Figure 5.7 presents the hierarchical cluster analysis of the upstream, downstream and sewage outfall SPE extracts combined with a heatmap. The heatmap represents all 1336 ions detected in all the DI-HRMS spectra. The dendrogram shows that all extraction replicates for each of

the sample groups cluster together. This confirms that the extraction replicates are more similar than the differences between the composition of the upstream, downstream and sewage outfall sampling sites. The hierarchical cluster analysis of the DI-HRMS spectra of the DOM extracts from Chew Stoke revealed that the composition of the sewage outfall and downstream were more similar, as these separate later in the dendrogram (Figure 3.18). However, at Betws-y-Coed the hierarchical cluster analysis of the DI-HRMS spectra of the DOM extracts found that the composition of the upstream and downstream were more similar, as these separate later in the dendrogram (Figure 5.7). So, the results of the hierarchical cluster analysis in both studies are consistent with the observations of the DI-HRMS spectra and the PCA.

The heatmap clearly shows those ions changing in abundance across the DI-HRMS spectra. Comparing the intensity of ions using the heatmap between the upstream and downstream DI-HRMS spectra shows that the overall composition of these extracts is very similar. Most ions are represented in similar colours indicating that the ion intensity in both the upstream and downstream DI-HRMS spectra are similar. However, there are clearly additional ions represented by warmer colours in the downstream which correspond to ions which are more abundant in the sewage outfall represented in red and blue in the upstream SPE extracts. This corresponds to what was seen in the DI-HRMS spectra in

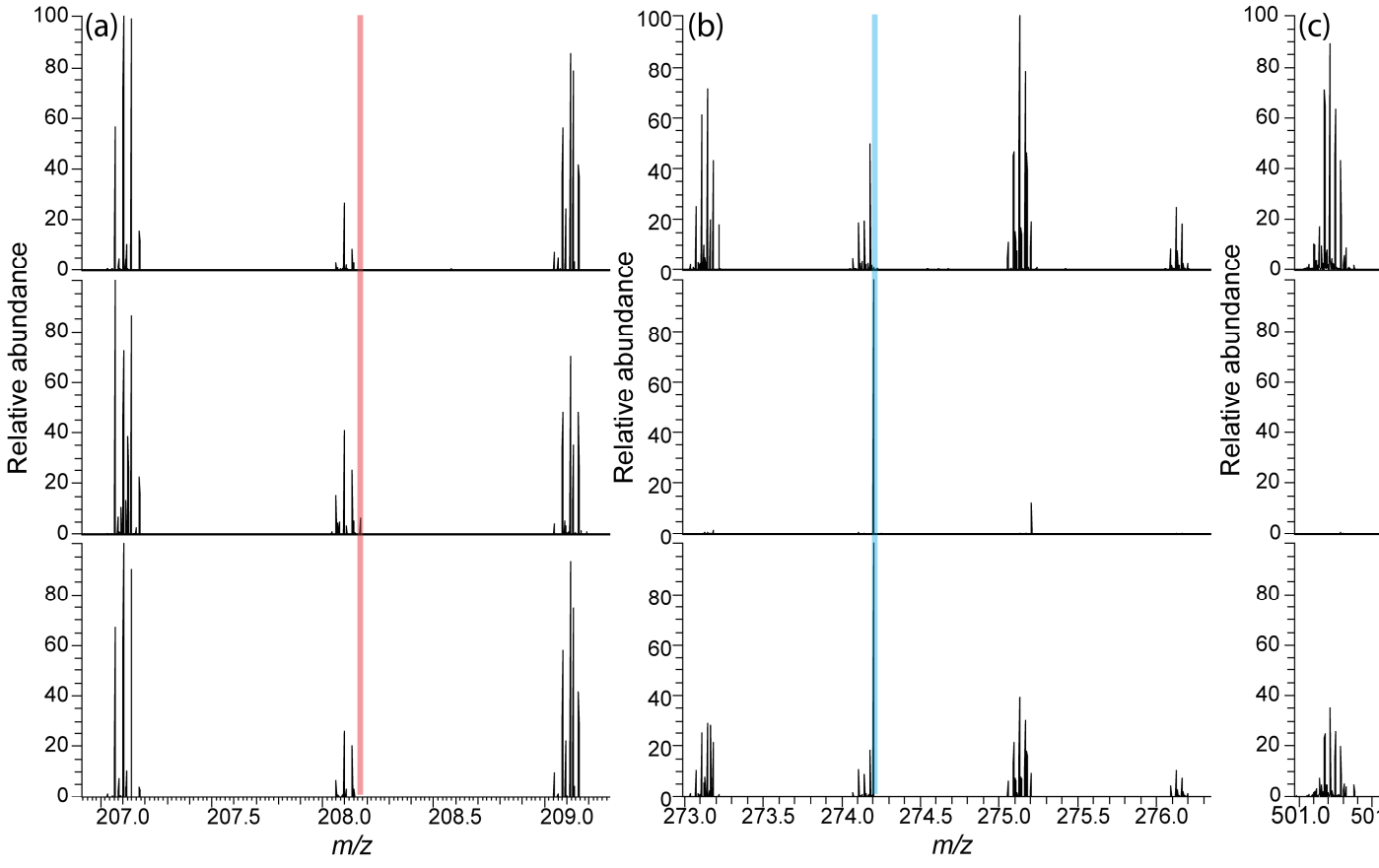


Figure **5.4** which showed additional ions in the downstream and sewage outfall DI-HRMS spectra which were not seen in the upstream SPE extract. Furthermore, there are ions which are in the upstream and downstream extracts represented in yellow or red which are comparatively represented in colder colours in the sewage outfall extracts.

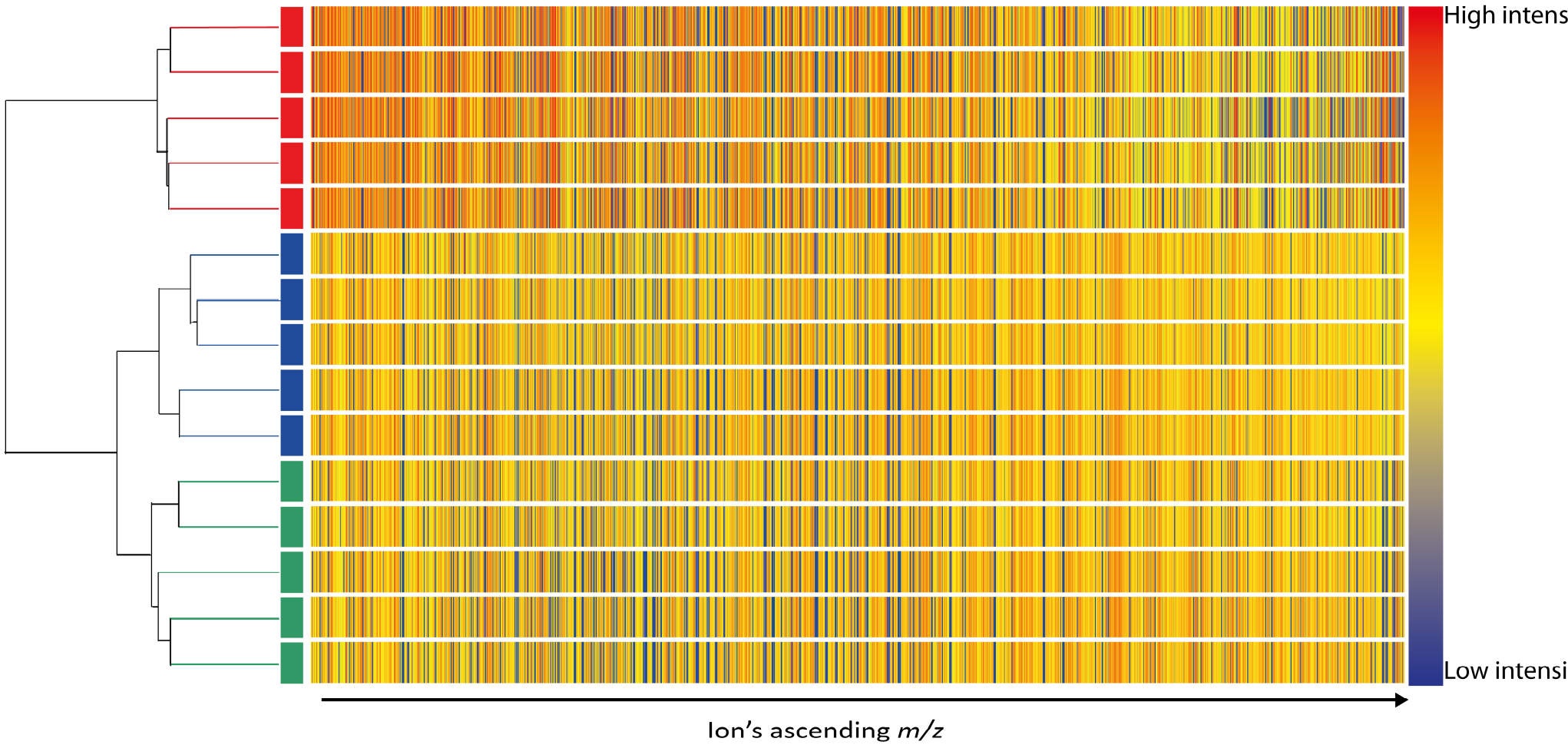


Figure 5.7. Hierarchical cluster analysis (left) of the upstream (■), downstream (■), sewage outfall (■) DI-HRMS spectra. Heatmap of the ions in in the DI-HRMS spectra arranges by mass (left to right). Comparison of the $\log_2$ of the intensity of the ions represented by colour with higher intensity hotter (red) and lower intensity colder (blue).

To determine which ions were most significant in differentiating between the composition of the DI-HRMS spectra of the riverine and wastewater DOM extracts, a Kruskal-Wallis analysis was used compare the upstream and sewage outfall DI-HRMS spectra. Table 5.2 shows the number of ions found at common p value significance thresholds. Furthermore, the ions are subdivided into two groups depending on whether an ion is increasing or decreasing in abundance in the sewage outfall DI-HRMS spectra compared to the upstream DI-HRMS spectra. At each of the p value threshold there are clearly more ions than predicted by chance. To assess which ions originated from the sewage outfall, only ions which were increasing between the sewage outfall and the upstream DI-HRMS spectra with a p value < 0.005 were selected for further analysis and identification. In total 337 ions were found to meet these two criteria.

Table 5.2. p values of ions from the Kruskal-Wallis analysis of the sewage outfall and upstream samples over a range of commonly used p value thresholds of statistical significance and the number of ions expected by chance.

p value	p all	p < 0.05	p < 0.02	p < 0.01	p < 0.005
Number of ions	1336	1030	959	898	559
Number of ions decreasing	632	466	451	336	222
Number of ions increasing	704	564	508	562	337
Expected by chance	N/A	66	26	13	6

5.4.2.3 Tentative identification of homologous series of ions

PPG was a significant component of the identified compounds originating from Chew Stoke sewage treatment works. PPG was characterised as a 58 Da spaced series of molecular ions in the DI-HRMS spectra in both the downstream and sewage outfall SPE extracts at Chew Stoke. In contrast the DI-HRMS spectra of the extracts from Betws-y-Coed showed no obvious series of molecular ions. To determine whether the same series of ions were present in the Betws-y-Coed DI-HRMS spectra, the pattern matching algorithm developed to extract series of ions at Chew Stoke was used. Three normally distributed series were found using this algorithm. These are shown in Figure 5.8 and summarised in Table 5.3.

From the accurate masses of the ions in each series the molecular formula of each ion was calculated using Equation 3.3. Of the three series identified in Betws-y-Coed sewage outfall, only one had a molecular formula similar to a series observed at Chew Stoke, i.e. Betws-y-

Coed's first sequence and Chew Stoke's fifth sequence. Of the 24 ions relating to PPG sequences all of them had a p value < 0.005. Finding series of ions with a 58 Da mass defect with a normal distribution of ion intensities, in both of the sewage outfall extracts indicate that this compound could be common across all sewage treatment works and used as a chemical marker.

Table 5.3. PPG sequences with their highest and lowest masses in the series, the range of the number theoretical monomer units, the range of p values for each series, molecular formulae of end groups, adducts and possible isotopes.

	Mass Range	**Monomer units**	**Minimum p value**	**Maximum p value**	**End Group and Adduct†**
PPG S1	171.0658-635.3995	2-10	0.0019	0.0031	NaO_2
PPG S2	243.0842-649.3787	2-9	0.0019	0.0031	$NaC_2H_4O_4$
PPG S3	257.1000-605.3524	3-9	0.0019	0.0025	$NaCO_3$

†Without further structural characterisation it is not possible to differentiate between the formula of the adduct or end group solely using exact masses. Therefore, the collective molecular formula of the adduct and end group is used in the last column.

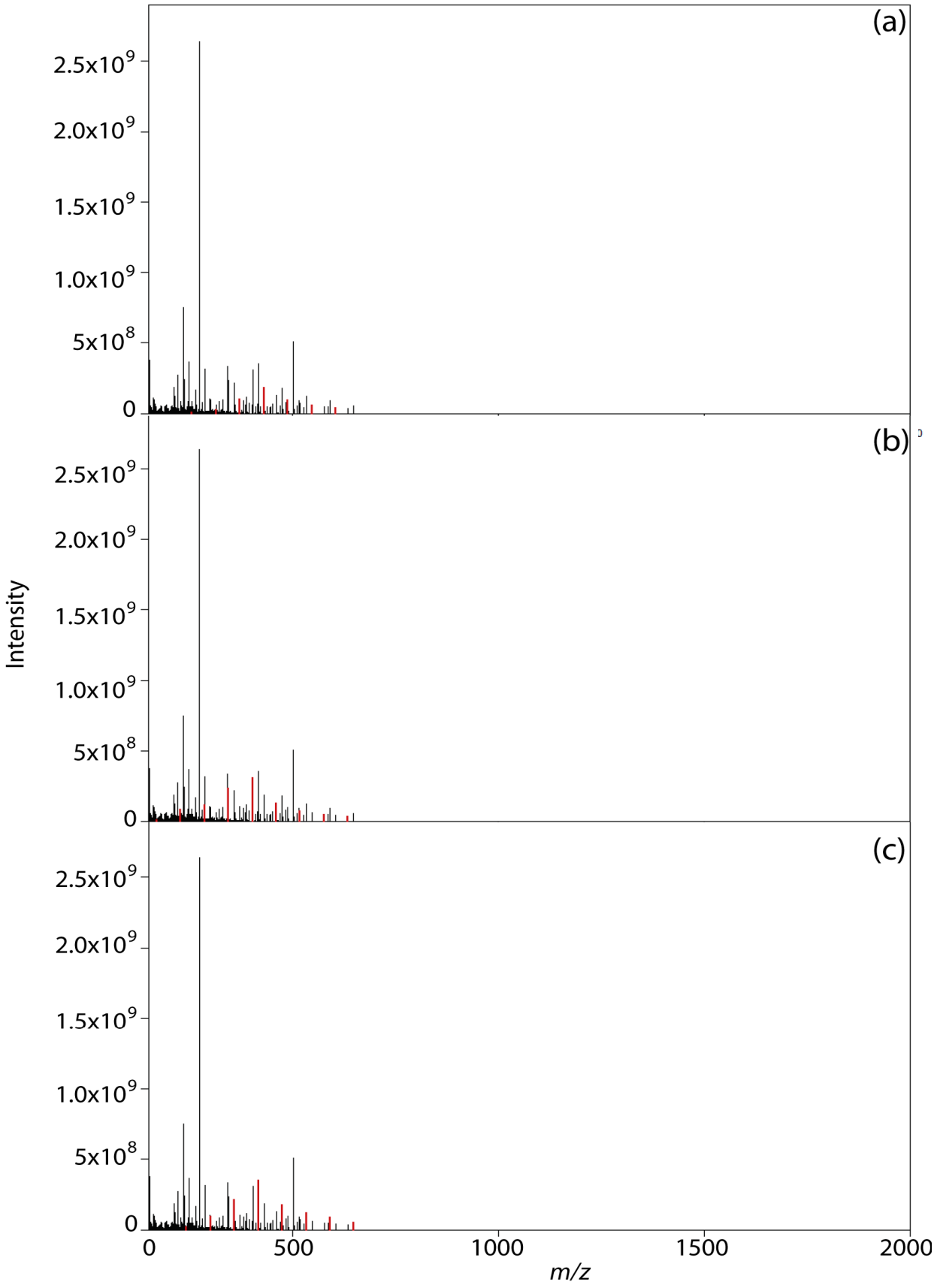


Figure 5.8. Sewage outfall direct infusion high resolution mass spectra from Betws-y-Coed with PPG sequences highlighted (red) (a) PPG sequence 1, (b) PPG sequence 2 and, (c) PPG sequence 3.

5.4.3 HPLC-HRMS of SPE extracts of DOM at Betws-y-Coed

One of each of the SPE extracts from Betws-y-Coed was analysed using HPLC-MS as described in Section 2.3.3. HPLC-MS analysis of the SPE extracts revealed further complexity of the DOM extracts. By observing the peak areas of individual EICs aligned across the analyses, a comparison between the extracts could be made. As in Chapter 4, the 337 ions found to be statistically significant in differentiating between the upstream and sewage outfall DOM DI-HRMS spectra from Betws-y-Coed were prioritised to be identified.

Figure 5.9 shows the TIC for the downstream, upstream, sewage outfall and blank extracts from Betws-y-Coed. There are several chromatographic features common across all chromatograms. From the TIC there is a clear difference between the sewage outfall SPE extract and the other extracts. However, due to extensive coelution of compounds meaningful comparison between the different extracts is impractical. The TICs of the blank, upstream and downstream SPE extracts are similar. However, Figure 5.10 shows that when using an EIC from the HPLC-MS analysis of a single ion compared across extracts, there is a peak in the upstream and downstream SPE extracts but not in the blank or sewage outfall extracts. This shows that there are compounds in the river not found in the blank or sewage outfall, not apparent when comparing the TICs.

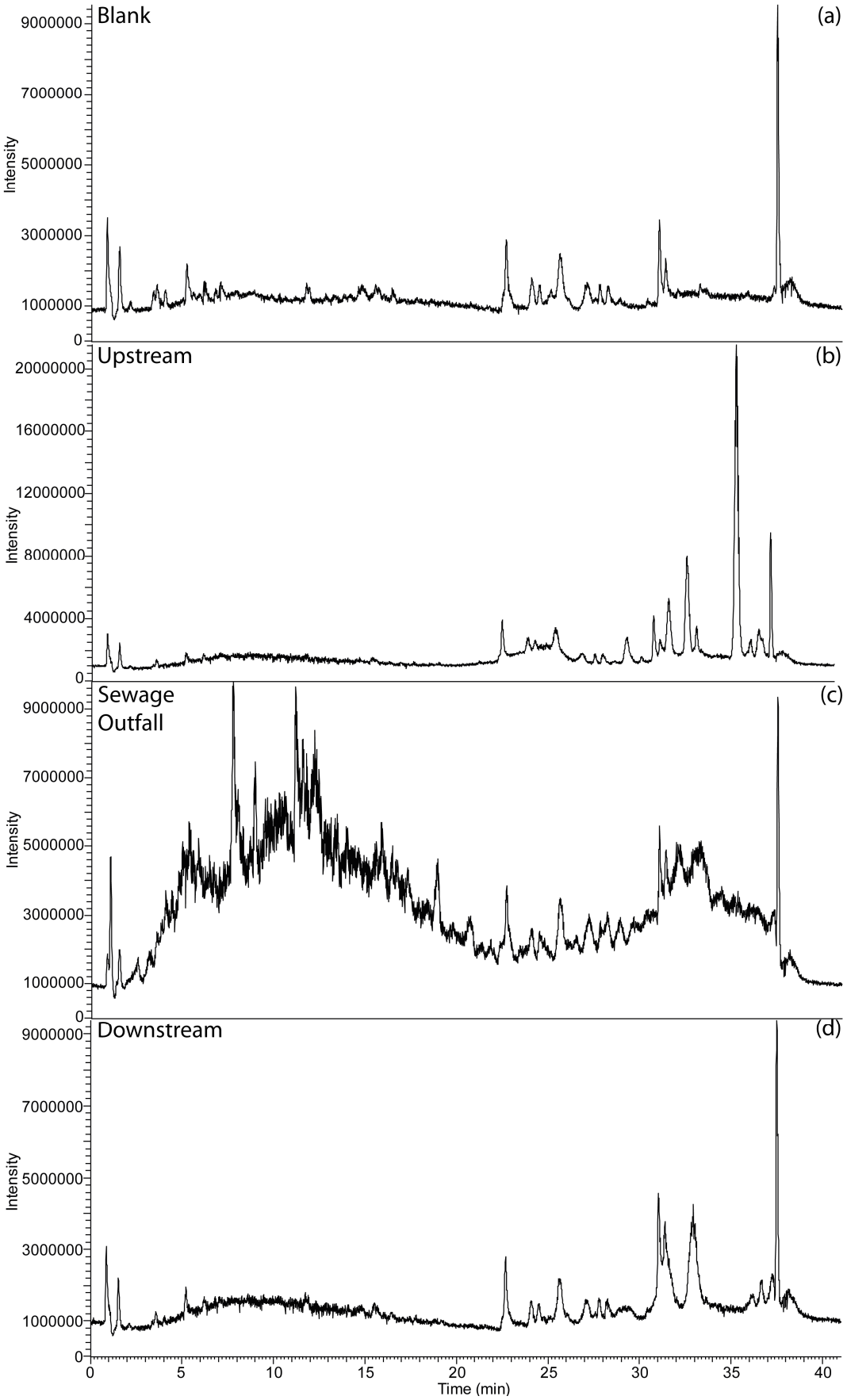


Figure 5.9. TIC of the blank, upstream, sewage outfall and downstream SPE water extracts analysed using HPLC-HRMS from Betws-y-Coed.

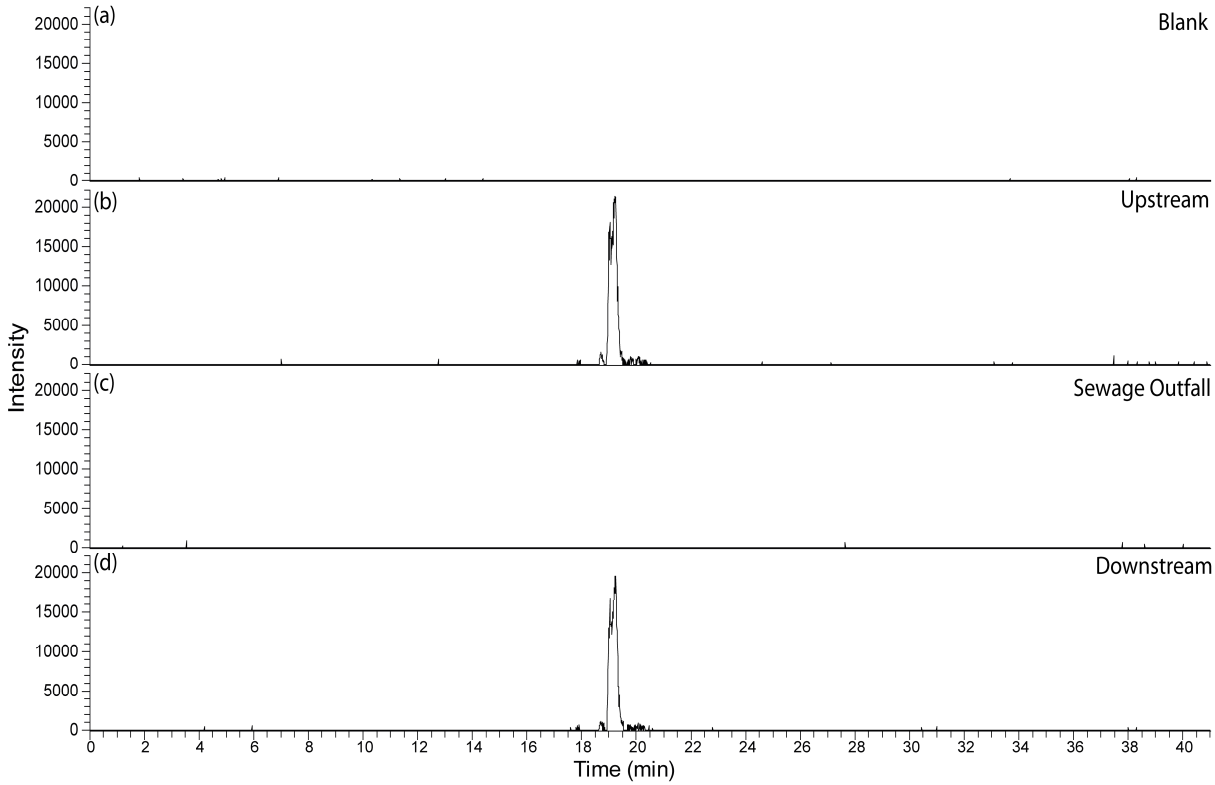


Figure 5.10. EIC for 10 ppm range of the *m/z* 400.3709 ion from the HPLC-HRMS in the blank extract, upstream, sewage outfall and downstream DOM extracts from Betws-y-Coed.

Figure 5.11 shows the UV chromatograms for the wavelength 254 nm all extract chromatograms there are peaks at 1 min and 37.5 min which have corresponding peaks in the TIC. In the blank there are no other features apart from an increase in absorbance which is correlated to the increasing concentration of acetonitrile in the HPLC gradient. By comparison, there is an area of increased absorbance between 2.5 to 18 minutes in the chromatograms of the riverine and sewage outfall DOM SPE extracts. The latter was also observed at Chew Stoke sewage treatment works in all of the DOM SPE extracts from the river and sewage outfall water. The measurement of 254 nm has been used across multiple different environments which has been correlated to the concentration of DOM in a system (De Haan and De Boer, 1987; Weishaar *et al.*, 2003; Yates *et al.*, 2016). Therefore, this may indicate there is a fraction of chromaphoric compounds which are present across all environments.

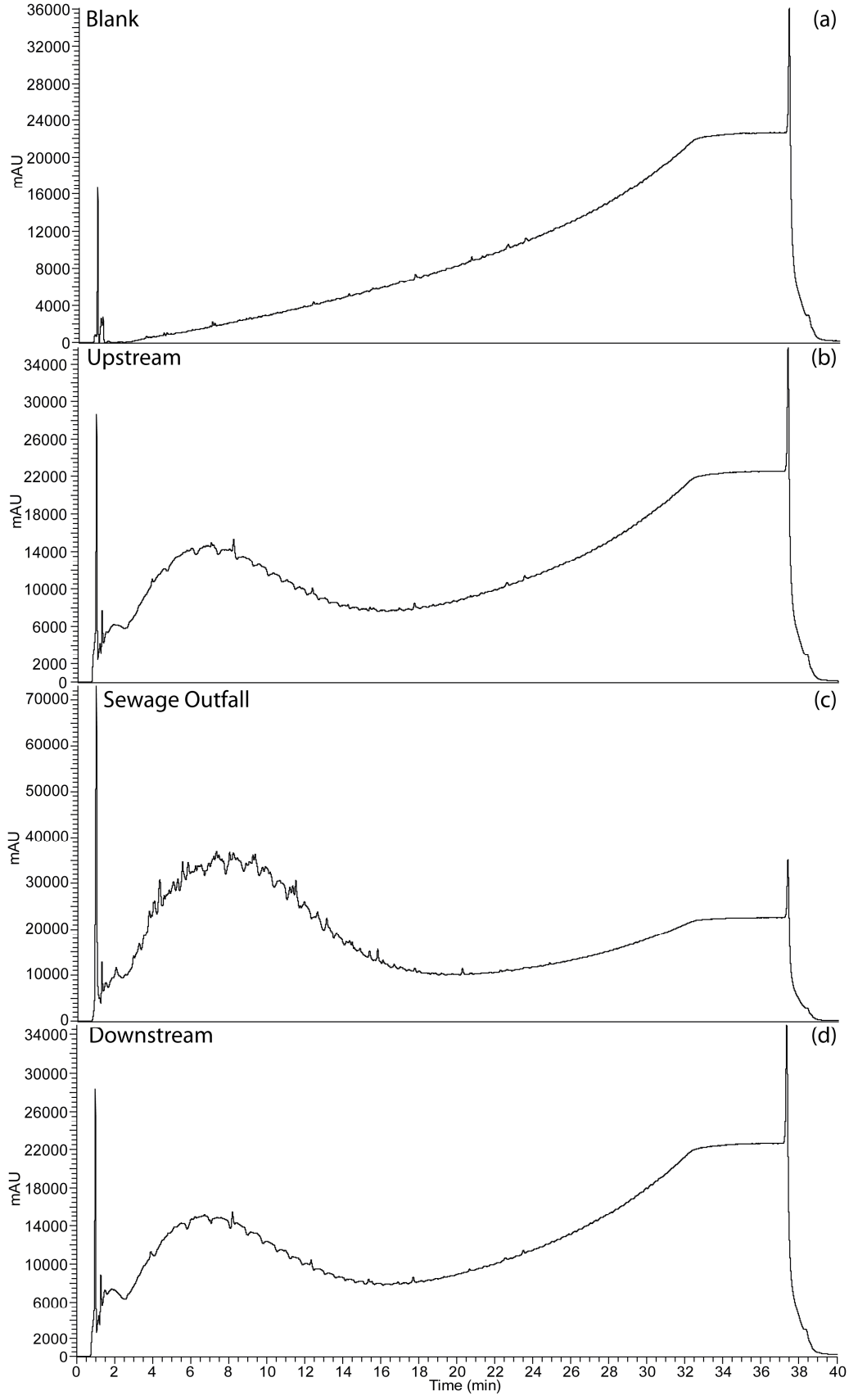


Figure 5.11 Specific UV absorbance chromatograms at 254 nm of the blank, upstream, downstream, sewage outfall SPE extracts from (top to bottom)

Peak picking and alignment as described in Section 2.3.3.1. After peak picking and alignment 9537 peaks were detected in all the Betws-y-Coed DOM and the blank extract. All peaks found in the blank were then removed leaving 7332 peaks aligned across all the samples. The ratio of the aligned the peak areas detected in the HPLC-MS analysis were then compared using a ternary plot (Figure 5.12). As at the Chew Stoke sewage treatment works, the ternary plot has 3 different sections highlighted.

Of the 7332 peaks detected in all the SPE extracts 5946 ions plot in green area which highlights ions likely originating from the sewage outfall. This shows that these ions have a larger peak area in the sewage outfall and downstream extracts and a lower or absent peak in the upstream extract. Ions plotting along the sewage outfall axis show correlations between the downstream and sewage outfall extracts. The majority of these ions plot at 100 % in the sewage outfall. This shows that the peak area proportional to concentration for these ions is higher in the sewage outfall HPLC-MS analysis than in the other two extracts. Some ions plot along the sewage outfall axis showing that these ions are also detected in the downstream and not the upstream extracts. However, the number of ions plotting along the axis, which would show an increasing peak area in the downstream, is much less than seen at the Chew Stoke site. This shows there are ions with a higher peak area in the sewage outfall but have a lower peak area in the downstream, consistent with higher dilution occurring by the Afon Llugwy than the River Chew. This is similar to the behaviour seen in the DI-HRMS spectra which showed that ions seen in the sewage outfall where reduced in abundance in downstream DOM extract DI-HRMS. To fully ascertain the diluent effect of the river the discharge rates of the sewage treatment works, and flow rates of the river would need to be measured.

798 ions plot along the downstream axis and are highlighted in blue. This indicates that these peaks are found in both the upstream and downstream extracts but were lower in abundance in the sewage outfall than the riverine DOM extract DI-HRMS, suggesting that these ions are riverine derived compounds. Furthermore, several ions plot towards 100% in the upstream DOM extract, likely indicating that the compounds giving rise to these are ions present in the upstream extracts are absent in the downstream extract; this is difficult to explain except by dilution or precipitation or interference within the MS.

Six hundred and forty-seven ions are situated in the centre of the plot, highlighted in red showing that these ions were found in all SPE extracts. This shows that part of the molecular

composition of all extracts is the same. This will be due to the water being used in sewage treatment process originating from local surface or ground water sources. In addition, the sewage is from the surrounding area where the water used will have come from local surface and ground water sources. Furthermore, there may be compounds already present in the river from other treatment works which are present in similar concentration to those in the sewage outfall.

No ions plot along the centre of the upstream axis which would indicate a correlation between the composition of upstream and sewage outfall SPE extracts. Instead, the ions plot towards origins of the axes indicating that these ions are either strongly correlated with the sewage outfall or the upstream extracts. Some ions plot at 100 % in the downstream, which indicates ions unique to the downstream which are not present in either of the other two extracts. This may be possible by either compounds generated by instream production or an additional source. The compounds found only in the upstream could either being used instream and therefore only found in the upstream extract.

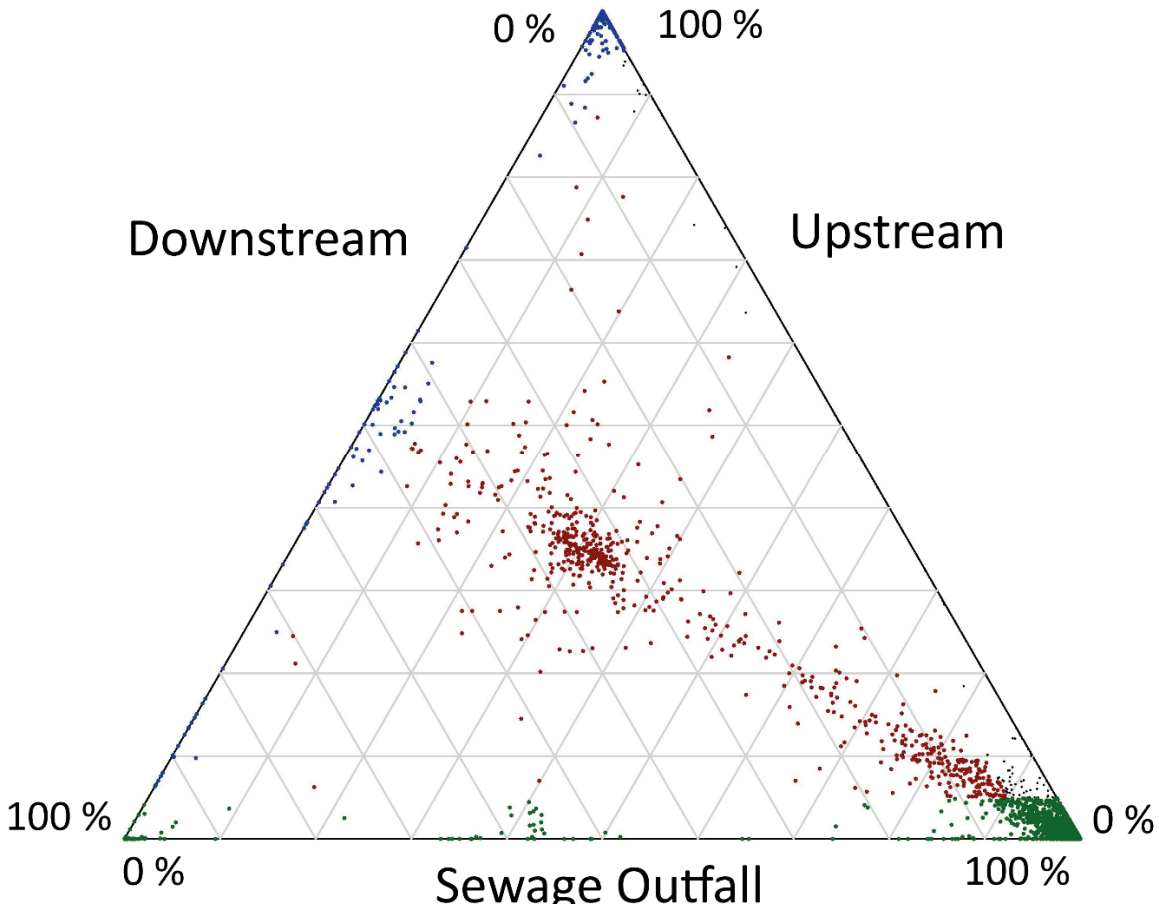


Figure 5.12. Ternary plot of ratios of the peak areas of the upstream, downstream and sewage outfall for all ions detected in the HPLC-HRMS analysis of the SPE extracts from Betws-y-Coed. < 5 % in the sewage outfall (blue), < 5 % in the upstream (green), > 5% in all samples (red).

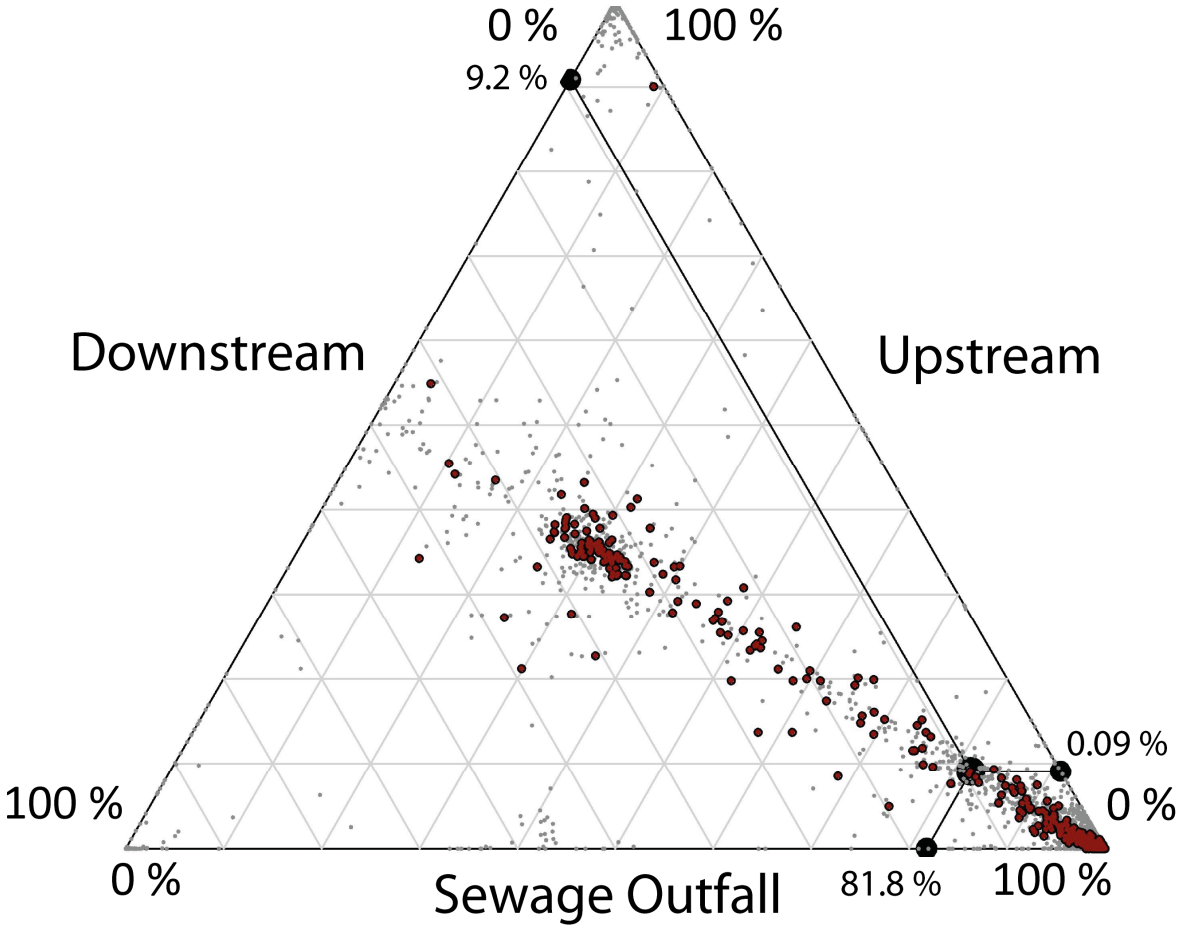


Figure 5.13. Ternary plot of ratios of the peak areas of the upstream, downstream and sewage outfall extracts for all ions detected in the HPLC-HRMS analysis (black), Ions highlighted with the same exact mass as the ions detected in the Kruskal-Wallis (red) The average of the ratios of ions with the same exact mass as highlighted by the Kruskal - Wallis plotted on each of the three axes.

The Kruskal-Wallis analysis of the DI-HRMS datasets in Section 5.4.2.2 found 337 ions with a p value < 0.005 indicating increased intensity in the sewage outfall extracts. Ions detected in the HPLC-MS within 10 ppm of the exact mass of the ions found to be significant using the Kruskal-Wallis analysis from the DI-HRMS were selected to be identified. From the HPLC-MS 549 peaks were found to have the same mass as 217 ions, highlighted in the DI-HRMS as being statistically significant. Figure 5.13 shows a ternary plot of the ratios of the peak's areas in the sewage outfall, upstream and downstream SPE extracts shown in black. The 549 peaks found to have the same exact mass as the 217 ions found by Kruskal-Wallis analysis to be statistically significant are highlighted in red. It appears that the peaks are evenly distributed between the centre of the plot and towards 100 % in the sewage outfall. However, the mean of the ratios of the 549 peaks found in the HPLC-MS analysis show that they have on average a

larger peak area in the sewage outfall extracts; 81.8 %. The downstream is 9.2 % which is approximately 10 times lower than the average in the sewage outfall and 10 times higher than the mean of the upstream. This clearly reflects the higher dilution of the sewage outfall by the Afon Llugwy when compared to the Chew Stoke sewage treatment works. HPLC-MS/MS for compound identification of significant ions originating from the sewage outfall at Betws-y-Coed

The 549 significant ions identified in the HPLC-HRMS data corresponded to the exact mass of 217 ions determined to be statistically by the Kruskal-Wallis analysis in the DI-HRMS analyses. These were compiled into a target mass list and a DDA carried out as described in Section 2.3.4. In a single run 2262 product ion spectra were recorded for 377 precursor ions. The product ion spectra were searched through the databases mzCloud, and MassBank.

5.4.3.1 Database results

As shown in Table 5.4 the product ion spectra of 5 compounds exactly matched to the database reference product ion spectra. Figure 5.14 shows the EIC of the 5 compounds identified. The sewage outfall clearly displayed the largest peak areas to the upstream and downstream SPE extracts, reflecting sewage outfall contribution and the dilution effects of the Afon Llugwy river flow. Figure 5.14(a) shows the EIC for the ion *m/z* 300.1592, which contain 6 peaks 3 of which are not fully resolved. For all of the peaks in the EIC the peak areas are largest in the sewage outfall extract. The product ion spectra of the peak labelled (ii) matched codeine, which was one of the 22 compounds identified in Chew Stoke sewage outfall. However, when comparing the EIC of Betws-y-Coed (Figure 5.14(a)) to that of Chew Stoke. Chew Stoke showed 3 clearly resolved peaks of which codeine was the largest. This shows how a single ion in the DI-HRMS spectra may comprise multiple structural isomers, and how variations in ion intensities in the DI-HRMS spectra may result from changes in concentration of multiple structural isomers.

The EIC of the ion *m/z* 212.1183 shows a single resolved peak. The product ion spectrum matched to the reference spectrum of diphenyl guanidine. This compound is used in the vulcanisation of rubber and in particular for the production of car tyres (Dvořák *et al.*, 2014). Diphenyl guanidine has been previously found in wastewater and is known to be acutely toxic to aquatic life, however, under standard ecotoxicity tests it was deemed to present little environmental risk (Canada Health & Canada Enviroment, 2013).

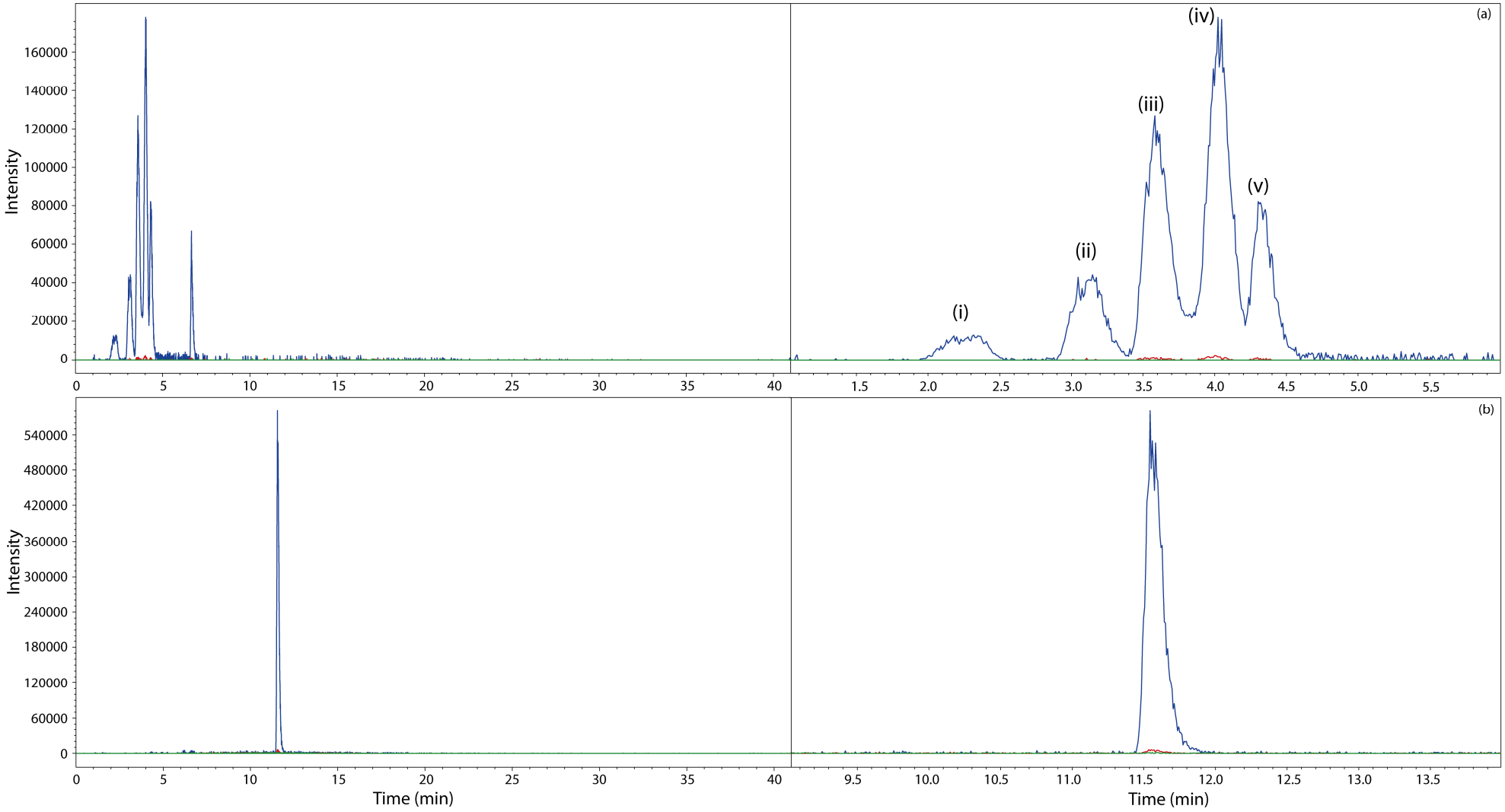


Figure 5.14. Overlaid EIC for 5 identified ions from the HPLC-MS of the sewage outfall (blue), downstream (red) and upstream (green) SPE extracts. Full EIC (left) 4 min time window around the identified chromatographic peak (right): (a) *m/z* 300.1562 to 300.1622 codeine peak ii (b) *m/z* 237.1000 to 237.1047 Carbamazepine

Table 5.4 Summary of the compounds identified in DOM from Betws-y-Coed.

Precursor ion (*m/z*)	Retention Time (min)	Fragmentation energy CID (eV)	p value	Molecular Formula	Compound	Database	Product ions (*m/z*)
300.1592	3.46	40	0.00193	$C_{18}H_{21}NO_3$	Codeine	mzCloud	282.1489, 267.1254, 253.1242, 243.1033, 225.0910, 215.1067, 199.0736, 193.0648, 187.0754, 183.0804, 175.0754, 165.0699, 161.0597
302.1754	3.68	30	0.0019305	$C_{18}H_{23}NO_3$	Dihydrocodeine	mzCloud	284.1657, 269.1416, 245.1180, 243.1032,241.1240, 227.1075, 217.1242, 213.0917, 201.0916, 199.0761, 195.0808, 189.0922, 185.0964
212.1183	5.62	40	0.00193	$C_{13}H_{13}N_3$	N,N'-Diphenyl guanidine	mzCloud	195.0923, 119.0607, 94.0653
237.1023	11.6	30	0.00193	$C_{15}H_{12}N_2O$	Carbamazepine	mzCloud	237.07, 220.08, 194.10, 192.08
389.1632	11.83	30	0.003058	$C_{21}H_{25}ClN_2O_3$	Cetirizine	mzCloud	201.0468, 187.1079, 186.0553, 166.0779, 165.0701

The EIC for the *m/z* 302.1754 ion showing two clearly resolved peaks. The earlier eluting peak's product ion spectra was identified as Dihydrocodeine, which is used in pain management. Codeine and dihydrocodeine both have the same pentacyclic structure, but their structures differ because of a double bond in the hydroxylated ring of codeine. Figure 5.15(a) and (b) compares the product ion spectra of codeine and dihydrocodeine. Both compounds produce similar product ion spectra. However, many of the fragments differ by 2 Da due to the double bond in codeine. Figure 5.15(c) and (d) show the product ion spectra corresponding to the peaks iii and iv from the EIC (Figure 5.14(a)). The precursor ion for both the product ion spectra have the same exact mass as codeine. Many of the product ions are similar to those seen in both codeine and dihydrocodeine which suggests that these compounds are structurally similar. There are many commonly known pentacyclic opiates including hydrocodone, heroine and morphine, however, they are readily distinguishable on the basis of their product ion spectra. Therefore, the compounds representing peak iii and iv may be transformation/degradation products or metabolites of one or more of compounds named previously which retain the pentacyclic structure.

The EIC of the ion *m/z* 237.1023 shows a single resolved peak Figure 5.14(b). This ion was identified as carbamazepine originating from the Betws-y-Coed sewage outfall and was also detected in the Chew Stoke sewage outfall. As discussed in Chapter 4 this is a commonly used pharmaceutical for the treatment of epilepsy and is acutely toxic to aquatic life (Miao *et al.*, 2005; Oetken *et al.*, 2005). Comparison of the EIC peak areas of carbamazepine at Chew Stoke with Betws-y-Coed shows the differences in the dilutional effects of the two rivers. At Chew Stoke the carbamazepine peak is clearly visible in the EIC whereas at Betws-y-Coed it is barely detectable in the downstream DOM extract (Figure 5.14(b)).

Cetirizine identified as one of the statistically significant compounds eluting from the sewage outfall at Betws-y-Coed. Cetirizine is a commonly used antihistamine employed in the treatment of hay fever and allergies. Unlike the other pharmaceuticals identified it is available as a non-prescription medication. It has been identified previously as originating from sewage outfalls but has limited ecotoxicity (Kosonen and Kronberg, 2009; Li, 2013; Almeida *et al.*, 2017).

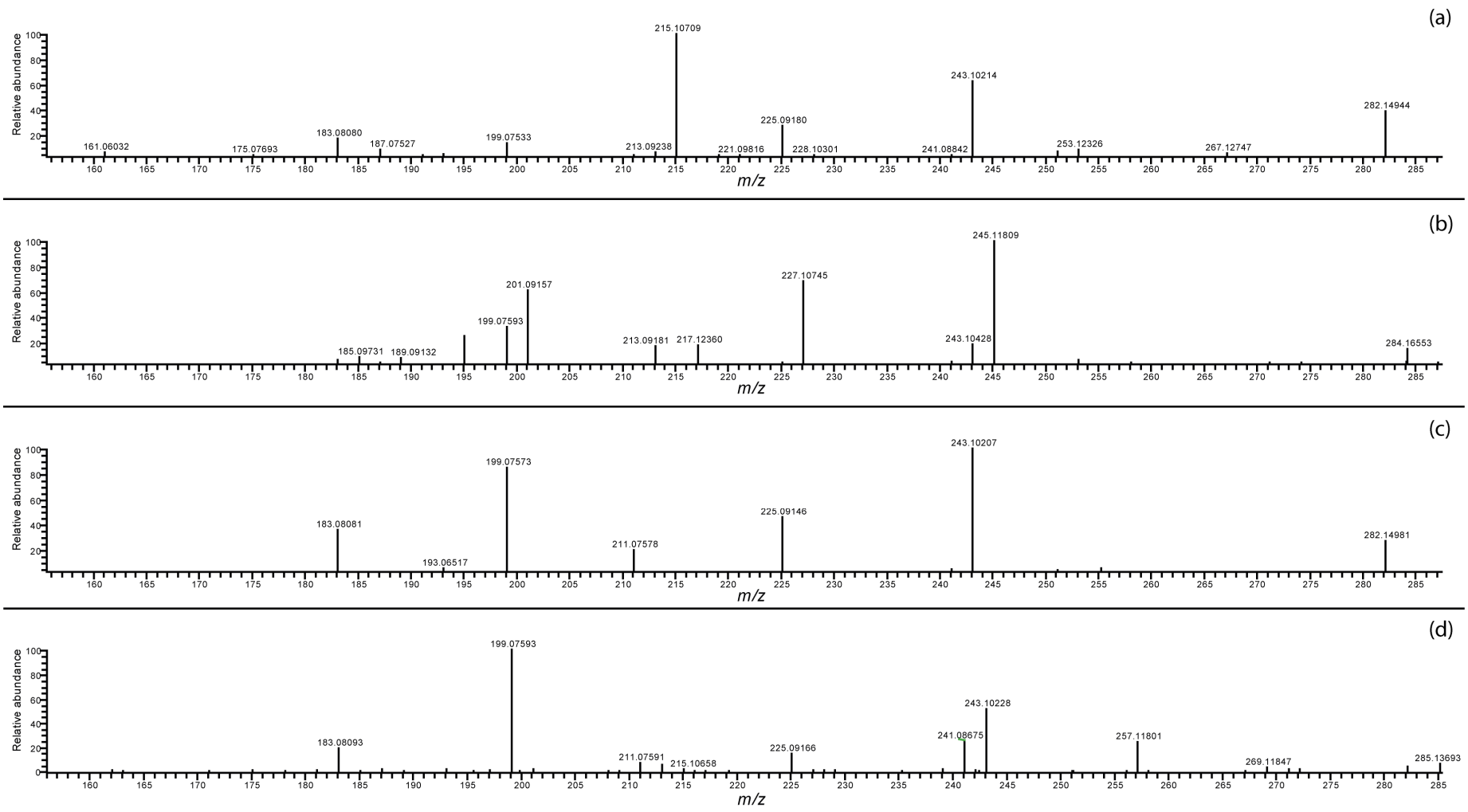


Figure 5.15. Recorded product ion spectra of: (a) Codeine *m/z* 300.1592 retention time 3.46 min CID energy 40 eV, (b) Dihydrocodeine *m/z* 302.174, retention time 3.68 min CID energy 40 eV, (c) *m/z* 300.1596 precursor ion corresponding to Figure 5.14.C peak iii ion retention time 3.98 min, CID energy 40 eV, (d) *m/z* 300.1594 precursor ion corresponding to Figure 5.14.C peak iv retention time 3.98 min CID energy 40 eV.

5.4.3.2 Polypropylene glycol identification

As shown in Section 5.4.2.3 three series of ions were tentatively identified as PPG from the DI-HRMS spectra of the sewage outfall extract of Betws-y-Coed (Table 5.3). It was found in the HPLC-MS analysis that all components of PPG series identified in the DI-HRMS spectra eluted from the column. Figure 5.16 shows an overlaid EICs of the ions corresponding to each of the PPG series identified from the DI-HRMS analysis. Figure 5.16(a) shows the EICs of the ions found to correspond to PPG series one of which was tentatively identified from the DI-HRMS spectra of the sewage outfall extracts. The peaks are poorly resolved on the column with most having an asymmetric peak shape, suggesting that they are co-eluting with other components of the same mass. Despite the poor resolution the chromatographic peaks form a normal distribution of peak areas indicative of oligomeric forms of a synthetic polymer.

Figure 5.17 shows the recorded product ion spectra of the precursor ion *m/z* 635.401 at 40 eV at the two different retention times corresponding to the two coeluting peaks at 17.48 and 18.00 min observed in Figure 5.16(a). Neither of the product ion spectra show series of 58 Da spaced product ions indicative of PPG as shown in Chapter 4. The major product ion for both product spectra in Figure 5.17 was *m/z* 563.378. This ion was also observed in the product ion spectra shown in Figure 4.17, 4.19, 4.22(b). Both product ion spectra are similar, however, the product ion spectra recorded at 17.48 min contains the ions *m/z* 515.32 and *m/z* 545.34 and the product ion spectra recorded at 18.01 min contains the ion *m/z* 521.32. Without the characteristic product ion spectra, which should show a series of 58 Da product ions, it is not possible to identify any of the latter components as PPG.

Figure 5.16(b) shows the EICs of the ions found to correspond to PPG series tentatively identified from the DI-HRMS spectra of the sewage outfall extracts. Clearly, each ion in the series produces two chromatographic peaks. The distribution of the peak areas for both series of peaks in the EIC show a clear normal distribution, again indicative of an oligomeric polymer. Figure 5.18 shows the two product ion spectra of precursor ions *m/z* 591.34 at 40 eV at the retention times corresponding to the two chromatographic peaks shown in the EIC in Figure 5.16(b). Neither product ion spectra show the clear 58 Da series of fragments which are used to characterise PPG series (Chapter 4). Therefore, without the characteristic product ions it cannot be confirmed whether these components are PPG.

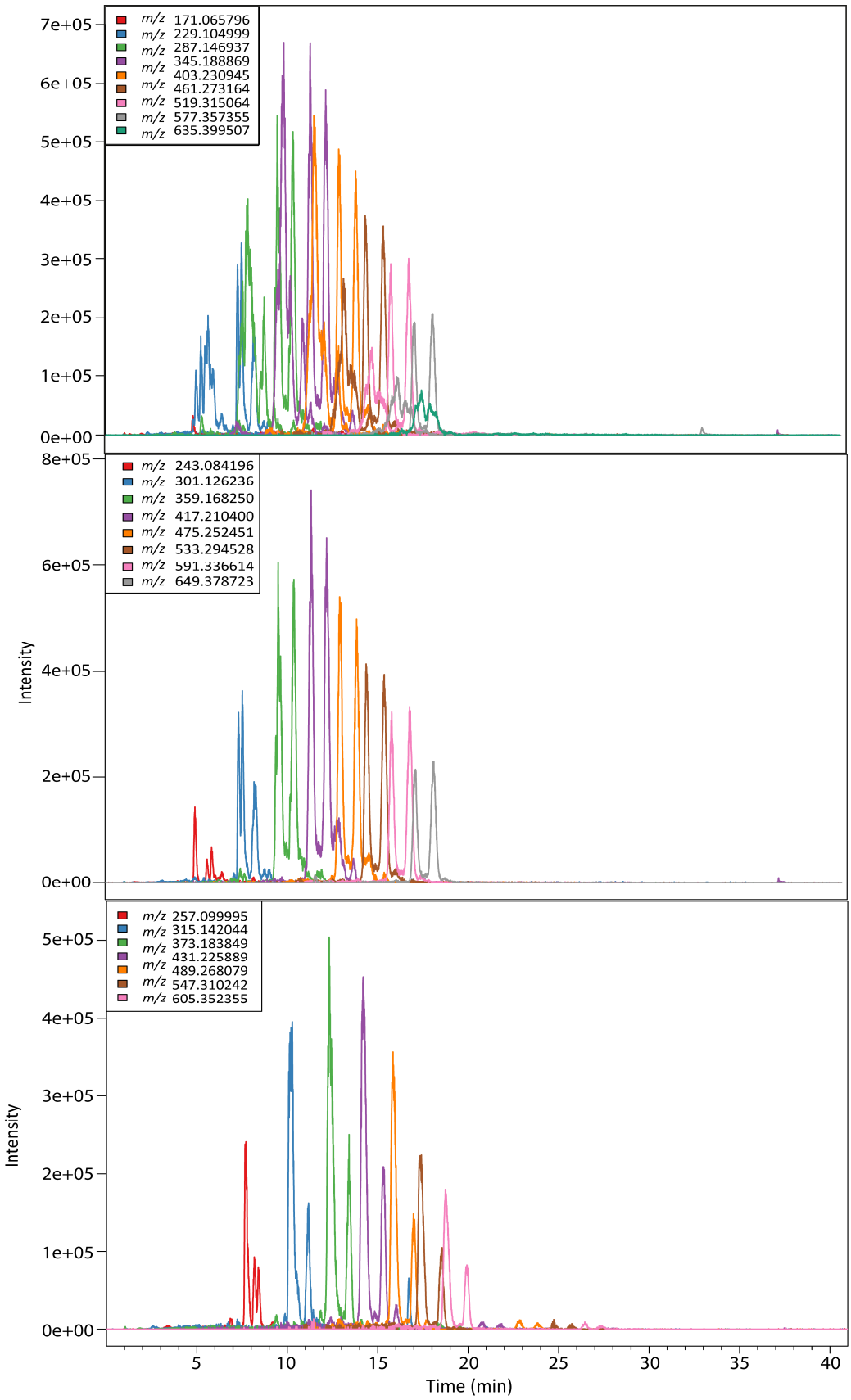


Figure 5.16. Overlaid EICs of the ions tentatively identified as PPG from the DI-HRMS spectra: (a) Ions corresponding to PPG series 1, (b) ions corresponding to PPG series 2, and (c) ions corresponding to PPG series 3.

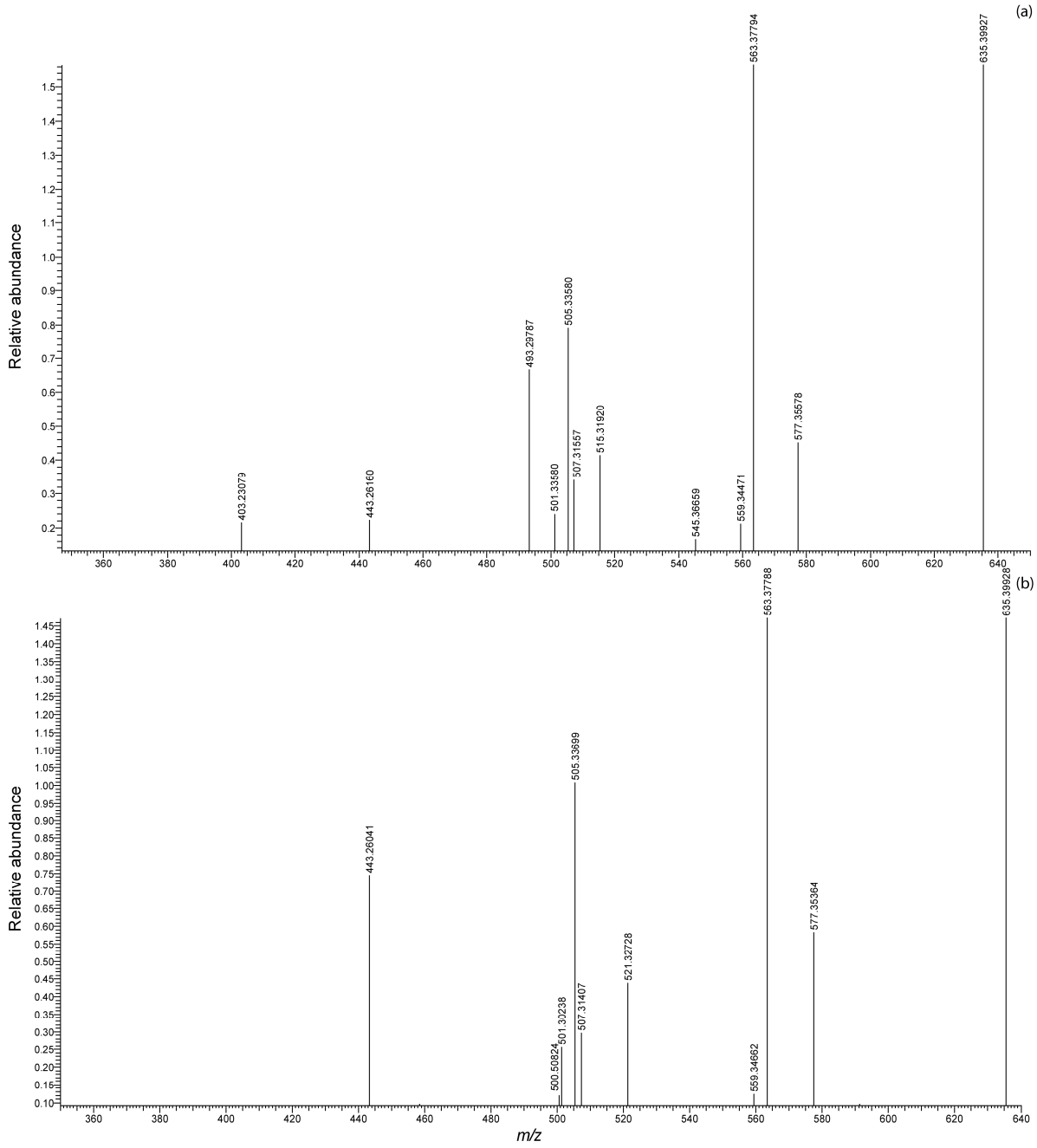


Figure 5.17. PPG series 1 product ion spectrum: (a) the precursor ion *m/z* 635.401, retention time 17.48 min, CID 40 eV, and (b) the precursor ion *m/z* 635.401, retention time 18.01 min, CID 40 eV.

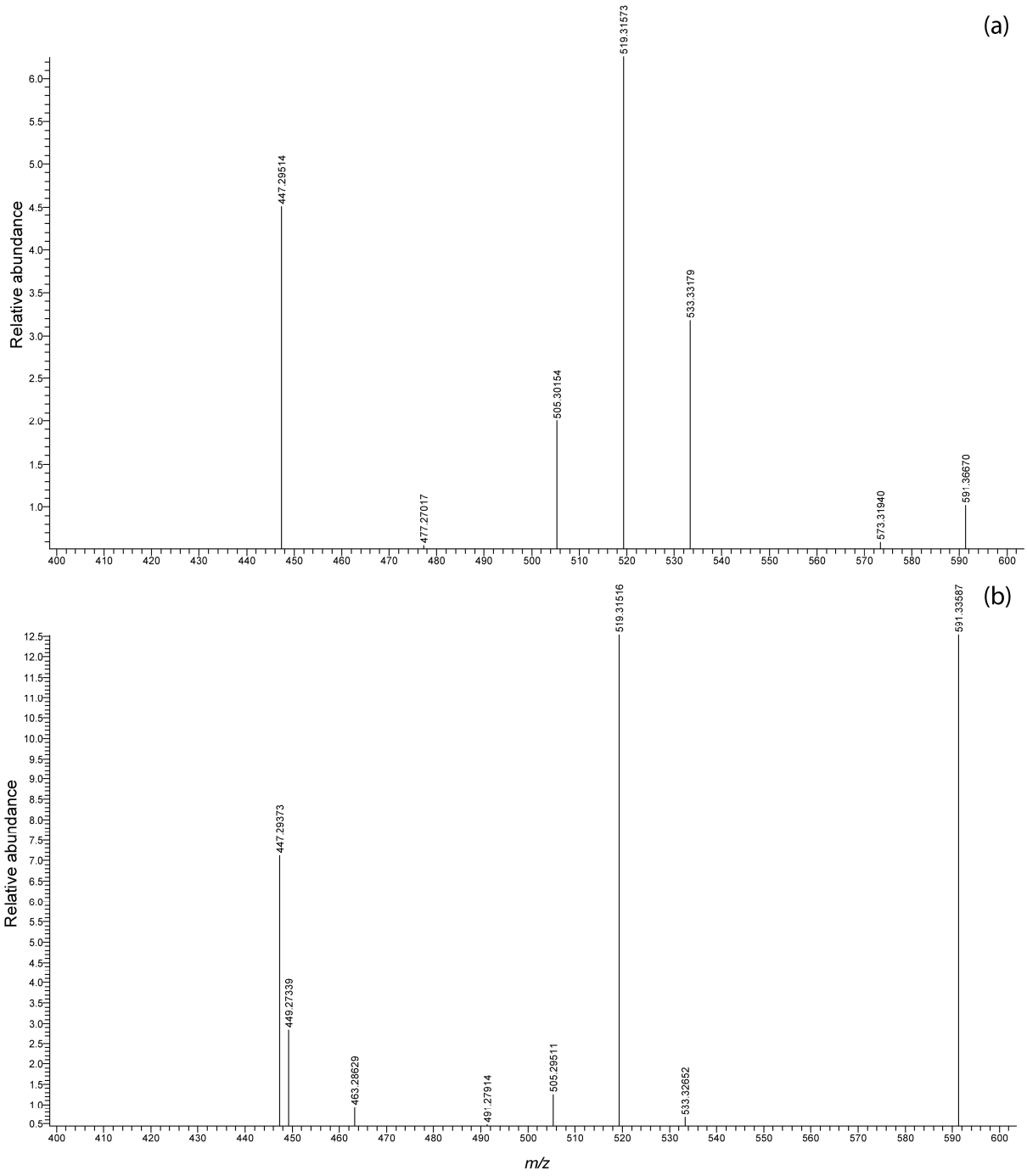


Figure 5.18. PPG series 2 product ion spectra: (a) the precursor ion *m/z* 591.337, retention time 15.85 min, CID 40 eV, and (b) the precursor ion *m/z* 591.337, retention time 16.80 min, CID 40 eV.

Figure 5.16(c) shows the EICs of the ions found to correspond to PPG series 3 tentatively identified from the DI-HRMS spectra of the sewage outfall extracts. Figure 5.19 shows two recorded product ion spectra of the two peaks in the *m/z* 605.352 EIC recorded at 40 eV (Figure 5.16(c)). Similar, to the other two series, the product ion spectra did not show the expected overlapping 58 Da spaced product ions. Therefore, neither series could be unequivocally identified as PPG.

This shows that even though the DI-HRMS spectra show a characteristic 58 Da spaced series with a normal distribution of intensities and the HPLC-HRMS analysis show a normal distribution of chromatographic peaks indicative of a synthetic polymer, the product ion spectra could not be used confirm these compounds as PPG, hence, they maybe correspond to a different polymer.

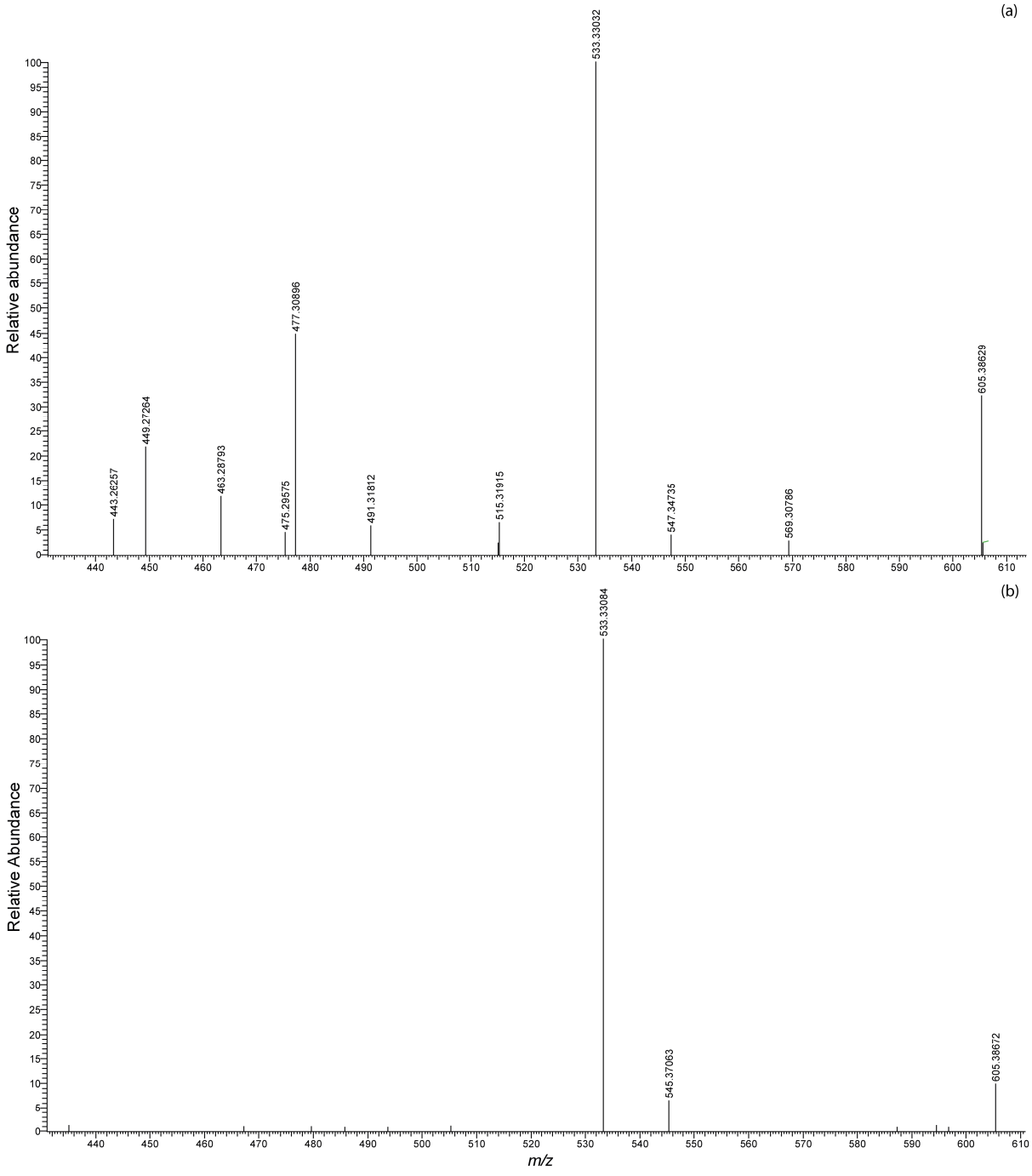


Figure 5.19. PPG series 3 product ion spectra: (a) the precursor ion *m/z* 605.352, retention time 18.60 min, CID 40 eV, and (b) the precursor ion *m/z* 605.352, retention time 19.70 min, CID 40 eV.

5.5 Summary

The sewage treatment works and riverine DOM extracts from Betws-y-Coed showed different distributions of ions in their DI-HRMS spectra (aim 1). Multivariate statistical analysis of the DI-HRMS spectra showed that the composition of the SPE extracts were significantly different (aim 2). The downstream spectra were shown to be an admixture of the upstream and sewage effluent DOM, which was not initially apparent from the manual assessment of the DI-HRMS spectra. Kruskal-Wallis analysis provided an effective data reductive step highlighting ions which were originating from the sewage outfall of which 5 were identified using subsequent HPLC-MS/MS analysis (aim 3 and 5). The Ternary plots of the components detected in the HPLC-MS analysis showed that the ions highlighted using the Kruskal-Wallis analysis were more abundant in the sewage effluent extract. Overall this showed that the method developed in the previous chapters has the potential to be applied to sewage treatment works in more complex environments.

By visually comparing the DI-HRMS spectra obtained from Chew Stoke and Betws-y-Coed the sewage outfall extracts differed in composition. This shows that it is unlikely that there is a "molecular fingerprint" for all sewage treatment works disproving hypothesis 1. This concurs with the results from previous targeted studies looking at the presence/absence and concentration of compounds in sewage outfalls, which was found to vary depending on the sewage treatment works sampled (Loos *et al.*, 2012). These differences in composition could be caused by a different: population demographics, product use, industrial activity in the surrounding area or the sewage treatment methods used.

By visually comparing the Conwy riverine extracts DI-HRMS spectra to that of the River Chew a common "molecular fingerprint" was not found which, disagrees with hypothesis 2. The differences in the riverine DOM could be due to the transportation mechanisms by which the water arrives at the river. As the water moves through different soil types and land uses different organic compounds will be deposited into the river, forming a different DOM profile for each river. By collecting a larger dataset of different river catchments and point sources it may be possible to determine if there are common fractions of DOM depending upon the catchment area and treatment works.

5.6 Conclusions

This chapter presents the results of using the methodology developed at chew stoke sewage treatment works (Chapters 3 and 4) to analyse SPE extracts from Betws-y-Coed sewage treatment works to answer the aims. The DI-HRMS, molecular "fingerprints" of the DOM from both riverine extracts and sewage effluent extracts differed. Although initial assessment suggested that the sewage effluent contribution was barely detectable in the river ca. 140 m below the sewage works, more detailed analyses showed that:

(i) By probing DOM composition at the molecular level resolvable through DI-Orbitrap-HRMS analyses it was possible to determine that the sewage effluent had affected the DOM composition downstream. (aim i)

(ii) PCA based on the DI-HRMS spectra of extraction replicates resulted in clustering into their respective sample groups. (aim ii)

(iii) PCA confirmed the upstream and sewage effluent DOM were least similar in composition and, as expected, the downstream DOM plotted between the two. (aim ii)

(iv) Hierarchical cluster analysis of the extraction replicates confirmed the similarities between the respective DOM sample groups. Furthermore, it showed that overall the downstream DI-HRMS spectra were more similar to the upstream spectra, than the sewage works spectra. (aim ii)

(v) Heatmaps provides an effective tool to visualise the differences and similarities in DI-HRMS spectra of the different DOM extracts. (aim i)

(vi) Comparison of the sewage outfall and upstream DOM DI-HRMS spectra using Kruskal-Wallis analysis provided a statistical data reduction step highlighting the 337 statistically significant ions which differentiate between the sewage effluent and riverine DOM. (aim iii)

(vii) Even though there was no obvious presence of an homologous series of ions with a 58 Da spacing (diagnostic of the presence of PPG) three were detectable using the pattern matching algorithm developed in Chapter 3.

(viii) HPLC-MS combined with peak picking emphasised the extreme complexity of the extracts revealing 7332 ions which were aligned across all three DOM SPE extracts.

(ix) Ternary plots provided a useful visualisation of the data acquired from the HPLC-MS analysis. Comparing the compositional differences between the different sites sampled. (aim iv)

(x) Of these ions, five compounds were identified including 4 pharmaceuticals and 1 plasticiser. All of which have been identified in previous targeted analyses of sewage effluents. (aim v)

(xi) The analysis of the series of ions with a 58 Da spacing was tentatively identified as PPG from the DI-HRMS analysis, when analysed using HPLC-MS/MS the product ion spectra for all series did not show fragment ions with a clear 58 Da difference as found in Chapter 4. Therefore, it was not possible to unequivocally identify these ions as PPG. (aim v)

Overall, this study showed that the methodology developed at Chew Stoke sewage treatment works was applicable to point sources in a catchment environment.

Therefore, it would be interesting to compare this study and Chew stoke with another comparable site. This will be explored in the next chapter.

Chapter 6

Untargeted analysis of compounds originating from Llanrwst sewage treatment works

Chapter 6. Untargeted analysis of compounds originating from Llanrwst sewage treatment works

6.1 Introduction

The previous studies investigated the sewage effluent from two treatment works with different treatment methods in different areas of the country. In both studies it was possible to differentiate between point source composition and the upstream riverine DOM. Furthermore, it was found in both rivers that the composition of the river downstream was an admixture of the DOM input from the point source and the upstream riverine DOM. Critically, however, it was found that the "molecular" fingerprint of both the river and sewage effluent DOM was markedly different. The geographical location and treatment methods used in the previous studies differ, a comparative treatment works which shares characteristics from both studies will form an interesting comparison.

Llanrwst sewage treatment works services the only town situated in the middle of the Conwy catchment. The sewage treatment works discharges into the main stem of the River Conwy. Figure 6.1 shows the position of Llanrwst sewage treatment works and highlights 460 km^2 sub-catchment of the river Conwy. A larger catchment area such as this will have a higher number of point sources discharging into the river. (Miao *et al.*, 2005; Kosonen and Kronberg, 2009; Karolak *et al.*, 2010; Golovko *et al.*, 2014; Ruff *et al.*, 2015). The compounds originated from these discharges are transported downstream in the river water which will make it more difficult to determine contributions to DOM from sources along the river.

Llanrwst and Betws-y-Coed are situated in the same catchment area and the Afon Llugwy is one of the three main tributaries forming the river Conwy. It will be interesting to compare whether the composition DOM in the River Conwy reflects that of the tributary at Betws-y-Coed. Llanrwst sewage works also provides a good comparison to Chew Stoke because it uses the same preliminary, primary and secondary treatment methods. Due to the river Conwy being larger than the Afon Llugwy and the River Chew this will test the sensitivity of this untargeted method in distinguishing source and riverine DOM.

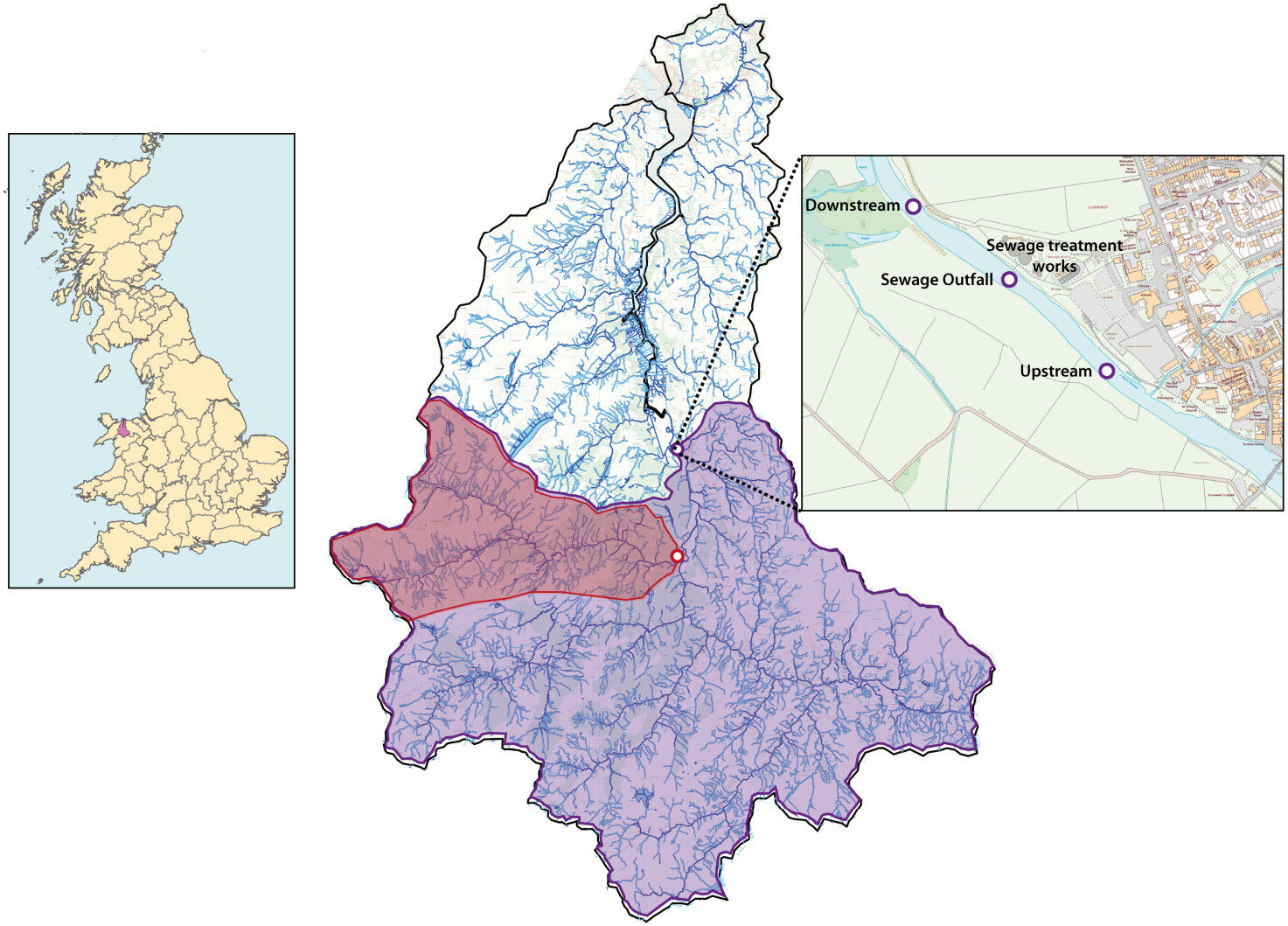


Figure 6.1. Map of the UK with the Conwy catchment area highlighted in pink. Map of the Conwy catchment area and river network. Red highlighted the Afon Llugwy sub-catchment upstream of the Betws-y-Coed sewage treatment works (O). Purple highlighted area is the sub-catchment of the River Conwy upstream of Llanrwst sewage treatment works (O). Smaller map showing the upstream, sewage outfall, and sewage outfall sampling sites around Llanrwst sewage treatment works

6.2 Aims

The overall aim of this study is to determine if the methodology developed at Chew Stoke and Betws-y-Coed can be used to investigate Llanrwst sewage treatment works DOM and compare and contrast the results from the three sites studied in this thesis.

The specific aims of this chapter are to:

(i) Explore the differences in the high-resolution DI-HRMS spectra of DOM extracted from Llanrwst sewage treatment works and the main river Conwy

(ii) Use multivariate statistics to determine if the DOM composition of sewage effluent from Llanrwst is significantly different to the DOM composition in the River Conwy.

(iii) Based on (ii) determine which compounds discriminate between the composition of the point source and the riverine DOM.

(iv) Identify the compounds determined to be statistically significant being originating from the sewage treatment works based on the results of (iii)

It is hypothesised that *the composition of the sewage effluent DOM from the treatment works at Llanrwst will differ significantly from the riverine DOM and will affect the composition of the river Conwy downstream of the treatment works*

6.3 Site information

Llanrwst sewage treatment works services discharges into the main stem of the River Conwy. Figure 6.1 shows the position of Llanrwst sewage treatment works and highlights 460 km^2 sub-catchment of the river Conwy. The water in the River Conwy is formed from three major tributaries; the Afon Machno, the Afon Lledr and Afon Llugwy. The previous study investigated the DOM composition of the sewage treatment works at Betws-y-Coed which discharges into the Afon Llugwy. The distance between Llanrwst and Betws-y-Coed sewage treatment works is approximately 5 km. Figure 6.1 compares the sub-catchments upstream of each of the sewage treatment works of the Afon Llugwy (110 km^2) and the River Conwy. This shows that even though these sewage treatment works are geographically close and within the

same catchment the sub-catchment upstream of Llanrwst covers an area approximately 4 times larger than sub-catchment of the Afon Llugwy.

The catchment upstream of Llanrwst drains from the peat bog situated on the Migneint uplands, which provides grazing land for sheep farming in the uplands and mixed livestock farming towards the lower half of the catchment. There are several villages upstream of Llanrwst including Betws-y-Coed, Capel Curig, Pentrefoelas, Dolwyddelan, and Ysbyty Ifan, which are situated on different tributaries. All these villages have dedicated sewage treatment works marked along the banks of the river, and Caple Curig has two sewage treatment works as discussed previously (Chapter 5).

Llanrwst sewage treatment works is situated on the banks of the River Conwy and the treated effluent is discharged into the river. Llanrwst uses comparable treatment methods to Chew Stoke. Sewage undergoes preliminary solid removal and primary settling. Its secondary treatment method uses filter beds for biological breakdown of organic matter but does not use a tertiary treatment method (Emmett *et al.*, 2016; Farkas *et al.*, 2018).

Llanrwst water samples were collected upstream and downstream of the sewage treatment works and directly from the water discharging from the sewage outfall. The upstream river water was collected approximately 217 m from the sewage treatment outfall. The downstream river water was collected 178 m downstream of the sewage treatment works.

6.4 Analysis of the composition of DOM extracts from Llanrwst

Sampling was carried out as described in Section 2.1. There was no visual difference between the water samples all were clear and colourless. The water samples and comparative HPLC grade water (Fischer Scientific) were filtered and extracted as described in Section 2.2 the extracts were fully dissolved in methanol/water (1:1, *v/v* ,2 ml). All extracts were yellow coloured, and there were no discernible differences between the three extracts. The blank extracts were clear and colourless.

6.4.1 DOC analysis at Llanrwst sewage treatment works

DOC analysis was carried out as described in Section 2.3.1 and the results are summarised in Table 6.1. The sewage outfall water has a higher concentration of DOM than the upstream or downstream water samples. Unexpectedly the concentration of DOM in the upstream water was more concentrated than the downstream water; differing by 1.15 mgL^{-1}, the reason for this

is unclear. Overall, however, the concentration of DOM extracted from the downstream water is comparable to the DOM extracted from the upstream water. The variability between replicate extractions is low and comparable to that achieved in the other chapters.

Table 6.1. DOC of the water sampled from each site from Llanrwst, together with calculated SPE extraction efficiencies.

Samples	DOC filtered water (mgl^{-1})	DOC extracts (mg)	Extract efficiencies (%)
Upstream	6.70 ± 0.001	2.81 ± 0.07	42.00 ± 1.60
Sewage Outfall	7.26 ± 0.04	3.82 ± 0.01	52.66 ± 1.43
Downstream	5.55 ± 0.04	2.77 ± 0.08	49.90 ± 1.52
Blank	N/A	0.28 ± 0.02	N/A

6.4.2 DI-HRMS of SPE extracts collected from Llanrwst sewage treatment works

All of the replicate extractions from the sampling sites and the blank SPE extracts were analysed by DI-HRMS as described in Section 2.3.2. A DI-HRMS spectrum of each of the DOM extracts from the Llanrwst sites comparing the full mass range *m/z* 150 to 2000 are shown in Figure 6.2. The DI-HRMS spectra show a clear compositional difference between the riverine and sewage outfall DOM. The upstream DOM DI-HRMS spectra show clusters of isobaric ions at 1 Da intervals with an overall normal distribution of intensities spanning *m/z* 150 to 800. This upstream DI-HRMS spectrum appears similar to the upstream DI-HRMS spectrum from Chew Stoke and previously reported DI-HRMS spectra of DOM (Flerus *et al.*, 2011; Osborne *et al.*, 2013; Mangal *et al.*, 2016).

The sewage outfall DI-HRMS spectrum of the SPE extract shows a series of ions with a consistent mass difference indicative of an homologous series of compounds. This is similar to DI-HRMS spectra of the DOM of the sewage outfall from the Chew Stoke. When comparing the DI-HRMS spectra of the sewage outfall and upstream DOM to the downstream DI-HRMS spectrum, the downstream DOM DI-HRMS spectrum appears to have a similar form to the upstream spectrum, suggesting that the sewage DOM has had little effect on the Conwy River DOM downstream of the point source.

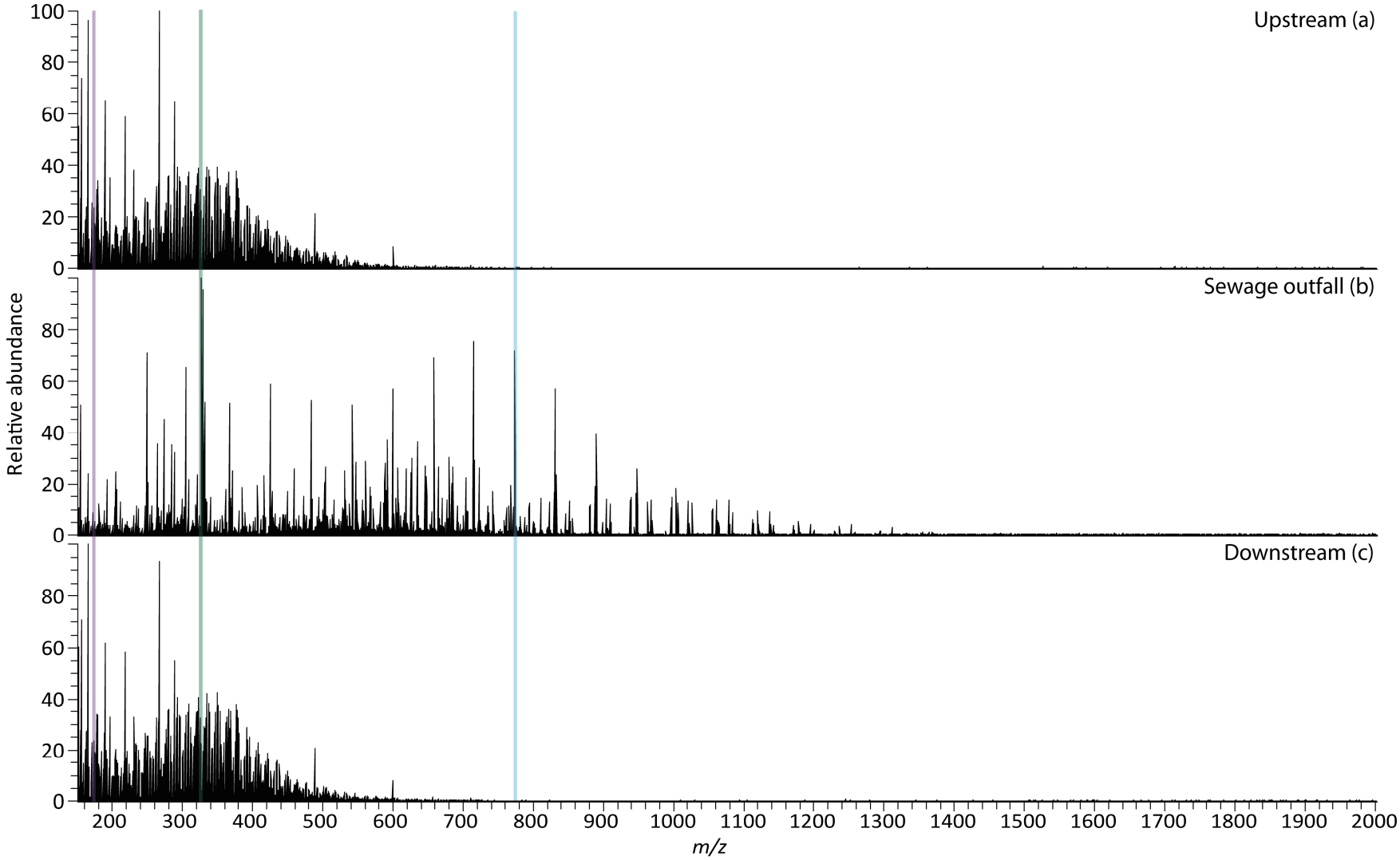


Figure 6.2. DI-HRMS spectra of SPE extracted DOM from the (a) upstream, (b) sewage outfall, and (c) downstream of the water samples collected from Llanrwst sewage treatment works, over the range *m/z* 150 to 2000.

Narrower mass ranges are compared in Figure 6.3 of the DI-HRMS spectra of the upstream, sewage outfall and downstream extracts. Figure 6.3 (a) compared the DI-HRMS spectra within the range of *m/z* 174.8 to 177.2, this range was chosen as no obvious differences between the three spectra could be seen in this mass range in the full mass range of the DI-HRMS spectra. Most of the ions are common in all the three spectra, however, there are two additional ions in the DI-HRMS spectra of the sewage outfall SPE extracts which are not seen in the downstream or upstream DOM extracts, *m/z* 175.1330 and *m/z* 177.1123 (highlighted in red). Furthermore, there are two ions in the DI-HRMS spectra of the downstream and upstream DOM extracts which are not seen in the sewage outfall DOM, i.e. *m/z* 175.0393 and *m/z* 177.0702. This shows that there are similarities in the composition of all three extracts, and there are also clear differences in the composition between the riverine and sewage effluent DOM extracts.

Figure 6.3(b) shows DI-HRMS spectra for the range *m/z* 326.0 to 329.4, which includes the most abundant ion in the sewage outfall DI-HRMS spectrum ion, *m/z* 327.0085(highlighted in blue). However, this ion is observed in all three DI-HRMS spectra. This was unexpected as the results from Chapter 3 and 5 showed that the most abundant ions in the sewage outfall DI-HRMS spectra were also found in the downstream spectra and absent from the upstream spectra. A possible explanation for this ion's presence in all three spectra is upstream of Llanrwst there are several small villages each with its own sewage treatment works, therefore, this compound may have been discharged into the river upstream of Llanrwst by a different sewage treatment works. However, when comparing the abundance of the ion at *m/z* 327.0085 between the three DI-HRMS spectra, the abundance is highest in the sewage outfall, lowest in the upstream and between the two in in the downstream DI-HRMS spectra. This increase in abundance in the downstream spectrum may be caused by the sewage outfall contribution.

Figure 6.3(c) compares the mass range *m/z* 773.0 to 776.0 which includes the ion *m/z* 773.5619. This ion is the most abundant ion of a series of ions spanning the higher mass range of the DI-HRMS spectrum of the sewage outfall DOM extract. The DI-HRMS spectrum of the sewage outfall extract contains three ions which have a mass difference of 1 Da and a decreasing abundance consistent with the ions *m/z* 774.5651 and *m/z* 775.5670 being the ^{13}C isotopes of *m/z* 773.5619 (highlighted in green). However, none of these ions are seen in the DI-HRMS spectra of the upstream or downstream SPE extracts. Instead broad asymmetric peaks are seen in both the downstream and upstream DI-HRMS spectra which indicates that these are most likely noise.

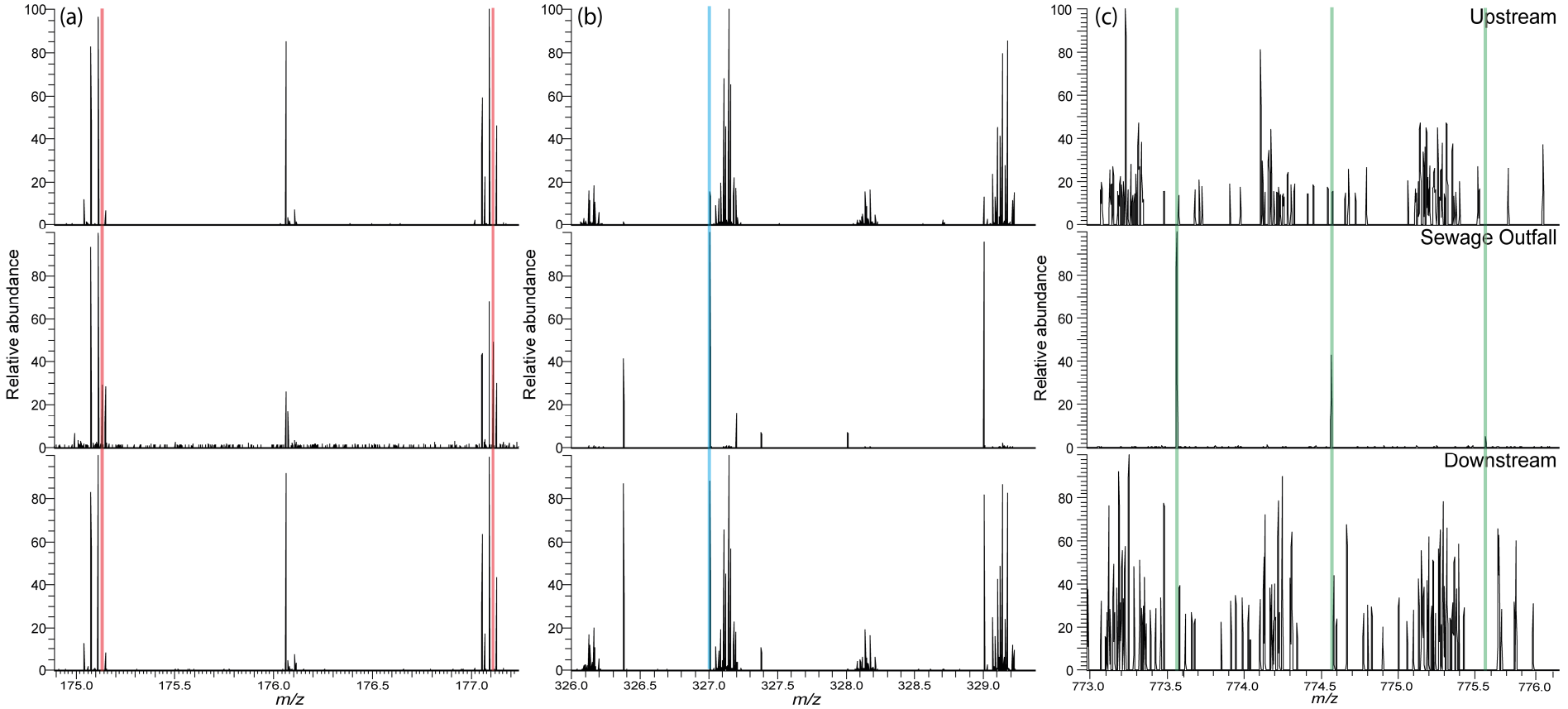


Figure 6.3. Comparison of narrower mass ranges of the DI-HRMS spectra of the upstream, downstream and sewage outfall SPE extracts extracted from Llanrwst at three different mass ranges. (a) Mass range *m/z* 174.8 to 177.2, (b) mass range *m/z* 326.0 to 329.4, (c) mass range *m/z* 773.0 to 776.0.

In summary, when comparing the full recorded mass range of the DI-HRMS spectra of the three DOM extracts from Llanrwst, there is a clear compositional difference between the sewage outfall and the riverine DOM extracts. Based on the examples above it appears that the sewage outfall DOM is barely detectable in the overall river Conwy DOM downstream. However, to determine if the sewage works has affected the Conwy DOM downstream PCA should show a difference between the upstream and downstream sampler as shown in Betws-y-Coed and Chew Stoke.

6.4.2.1 Ion picking and alignment

Ion picking and alignment of all the replicate downstream, upstream, sewage outfall, blank and QCs analysed using DI-HRMS was carried out as described in Section 2.3.2.1. The scan to scan variability of the instrument was 3.2 ppm for Llanrwst. The mass accuracy of standard ions over the analyses had a range of 2.12 ppm. Therefore, the ions were aligned using their exact mass using a 5 ppm mass range. Across all DI-HRMS spectra, 3393 ions were picked and aligned for all SPE extracts of water collected from Llanrwst.

6.4.2.2 Llanrwst Statistical analysis

PCA was used to compare the composition of the blank, upstream, downstream, sewage outfall and pooled QCs from Llanrwst (Figure 6.4). From the PCA, not all of the sample groups separate. There is clear overlap between the downstream and upstream replicates showing that these extracts are not significantly different. No further separation was observed in any of the other lower PCs. The blank is clearly separated from the QC, sewage outfall and riverine groups in PC1 which explains 39.84 % of the variance between the data sets. The pooled QC forms a cluster between the sewage outfall and riverine SPE extracts as expected. This shows that there were no large changes during analysis which could be affecting the results. The sewage outfall shows clear separation from the riverine extracts in PC2 which explains 31.12 % of the variance.

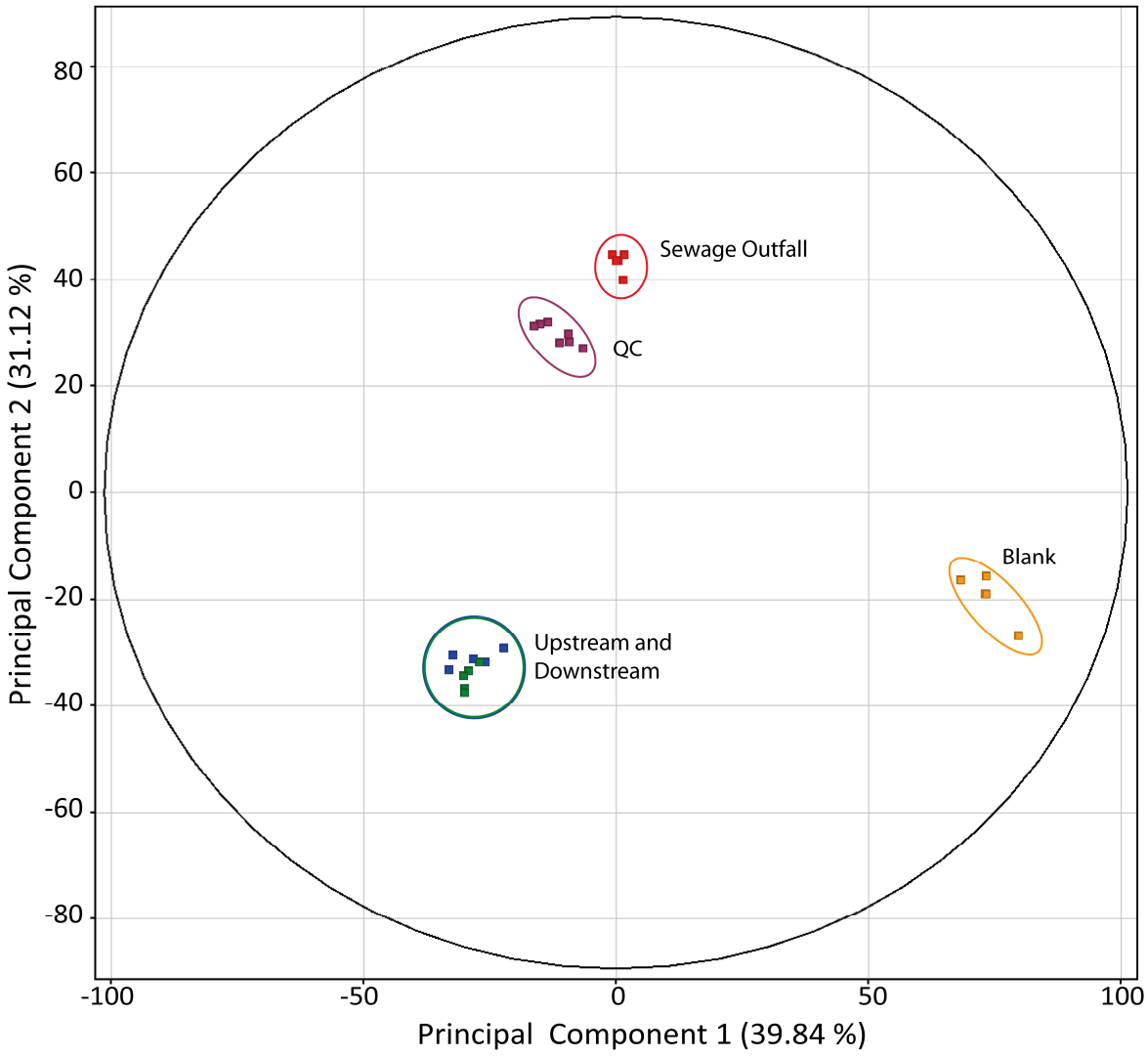


Figure 6.4. PCA analysis of the ions found in the Llanrwst DI-HRMS spectra of the upstream (■), downstream (■), sewage outfall (■), QC (■) and blank (■) SPE extracts.

PCA of the upstream, downstream and sewage outfall extracts is shown in Figure 6.4. The PCA shows a significant difference between the composition of the DI-HRMS spectra of the sewage outfall and riverine SPE extracts in PC1 which explains 64.15 % of the variance. Unlike at Chew Stoke and Betws-y-Coed, the composition of the DOM extracted downstream of the sewage outfall is not significantly different to composition of the upstream SPE extract. This shows that the compounds coming from the sewage outfall are either: (i) already present in the river and the change in the concentration caused by the sewage outfall is insufficient to cause a significant change in composition to differentiate between the DOM extracts, or (ii) the sewage effluent is so diluted by the river water and that the compounds essentially become below the limit of detection in the downstream DOM DI-HRMS analysis.

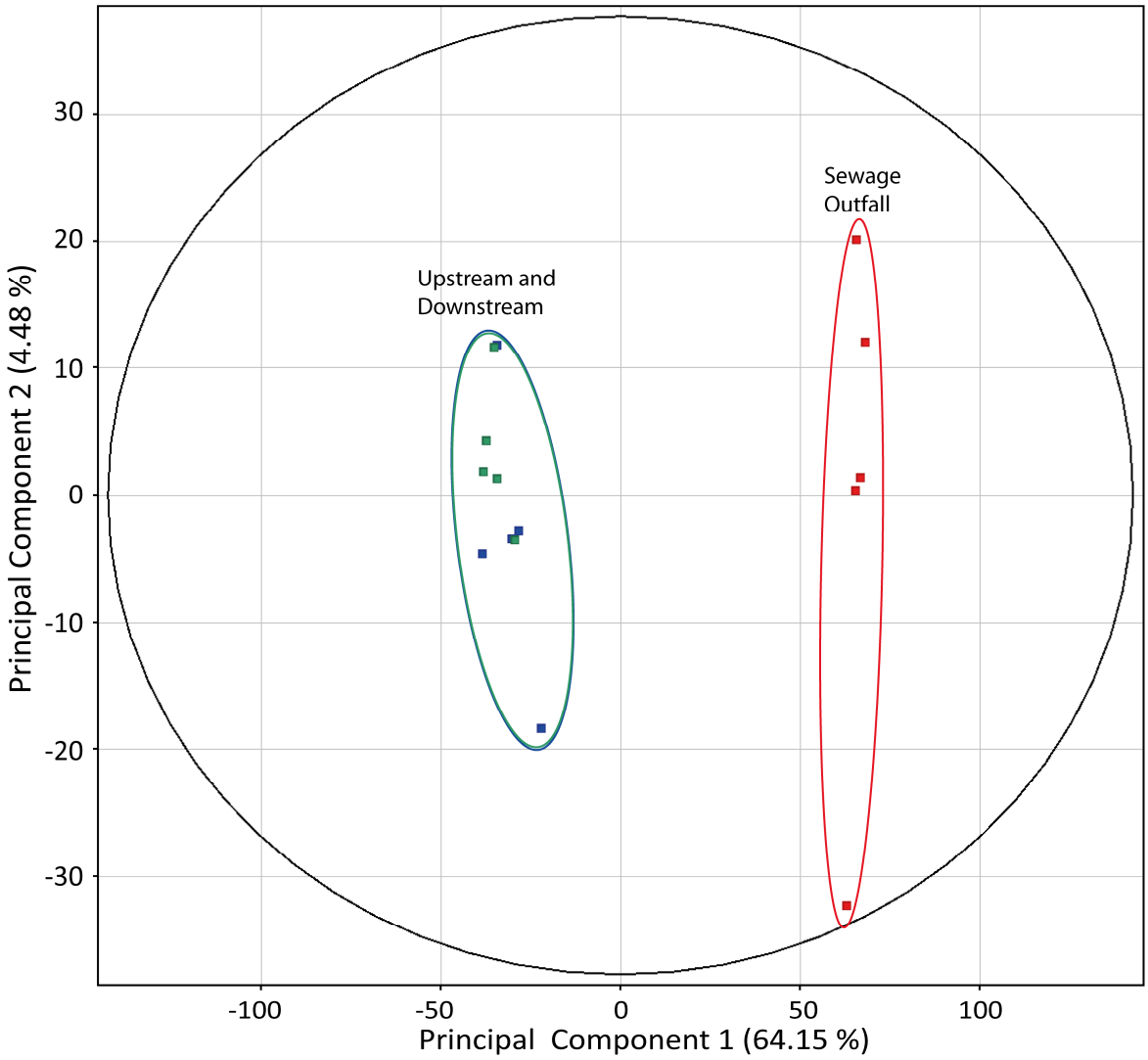


Figure 6.5. PCA analysis of Llanrwst of the DI-HRMS spectra of the DOM extracts from the upstream (■), downstream (■) and sewage outfall (■) water.

Figure 6.6 presents the hierarchical cluster and heatmap for the extracts collected from Llanrwst. The dendrogram shows that overall the sewage outfall DOM composition is least similar to the downstream and upstream compositions. The downstream SPE extracts are closely related to the upstream and do not completely separate on the dendrogram. One of the upstream SPE extraction replicates clusters with the downstream replicates. This shows that the composition of the riverine SPE extracts are not different. Comparing the DI-HRMS spectra using the heatmap show clear differences between the riverine and sewage outfall DI-HRMS spectra. Both riverine DOM extracts display similar ion intensities as represented in the similar colours. Furthermore, there appear to be no ions in the heatmap which are found in the sewage outfall and downstream replicates but not found in the upstream, which contrast with what was seen at Betws-y-Coed and Chew Stoke (Figure 5.7 and 3.19).

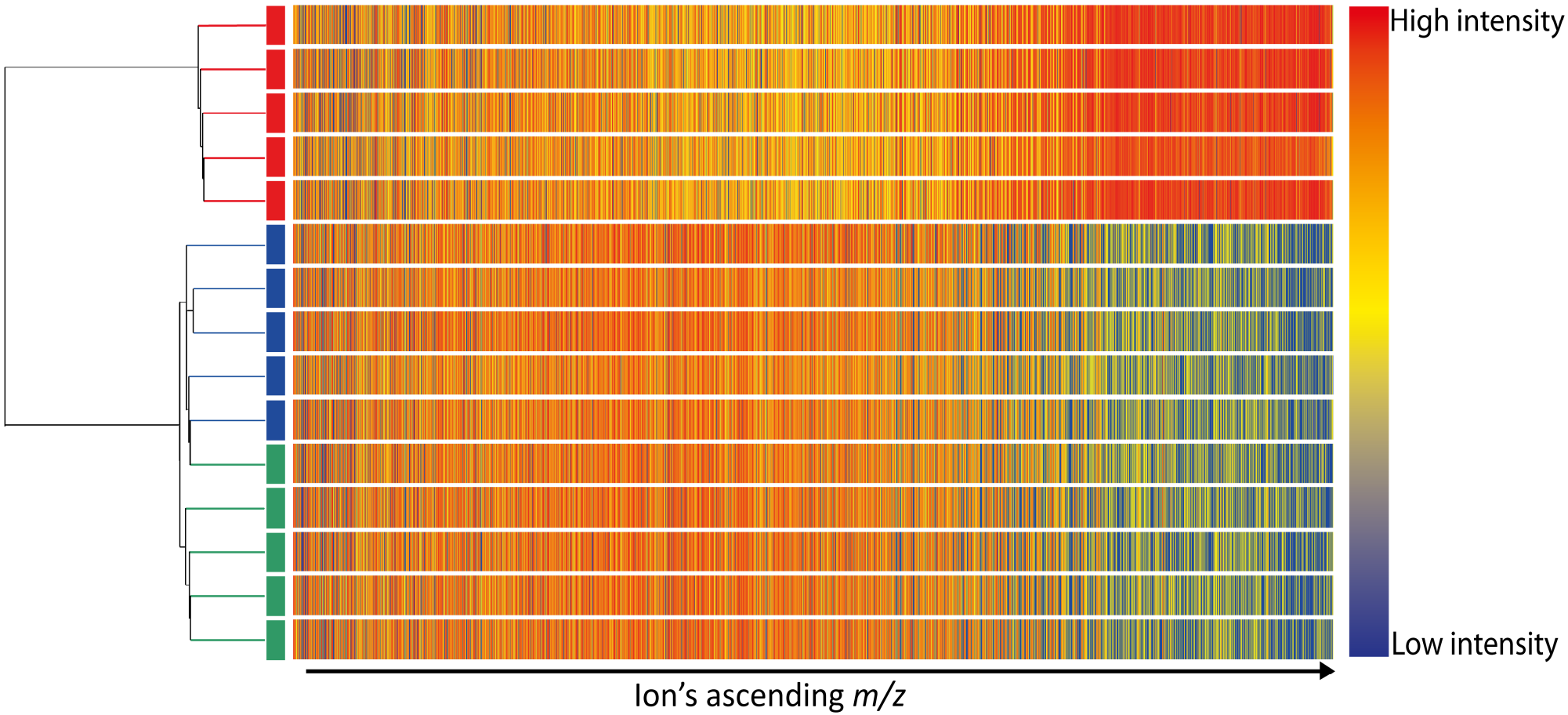


Figure 6.6. Llanrwst Hierarchical cluster analysis (left) of the upstream (■), downstream (■), sewage outfall (■). Heatmap (right) of the ions in in the DI-HRMS spectra arranges by mass (left to right). Comparison of the $\log_2$ of the intensity of the ions represented by colour with higher intensity hotter (red) and lower intensity colder (blue).

To determine which ions, differ significantly between the sewage outfall and upstream DOM DI-HRMS spectra, Kruskal-Wallis analysis was used to compare both spectra. Table 6.2 shows the number of ions decreasing and increasing in the DI-HRMS spectra of the sewage outfall extracts when compared to the upstream DOM extracts. The p values of the ions were compared to common significance thresholds. At each of the p value thresholds there are clearly more ions than predicted by chance. To assess which ions, originate from the sewage outfall, only ions which increased in abundance in the sewage outfall and had a p value <0.005 were selected for further analysis and identification. In total 174 ions were found which met these criteria.

When comparing the number of significant ions found in each of the sewage treatment works studies, there is a noticeable decrease in the number of significant ions with a p value <0.005 and increasing abundance in the sewage outfall extracts., In Chew Stoke 510 ions were found and in Betws-y-Coed 337 ions. It is difficult to compare these numbers directly as these all refer to DI-HRMS spectra from different studies with different numbers of ions found in each dataset. However, this could indicate that the difference between the riverine and sewage outfall DOM is decreasing as the complexity of overall river DOM increases due to mixing of multiple diffuse and point source inputs, including a number of source sewage works contributions.

Table 6.2. p values of ions from the Kruskal-Wallis analysis of the sewage outfall and upstream samples over a range of commonly used p value thresholds of statistical significance and the number of ions expected by chance.

p value	p all	p<0.05	p<0.02	p<0.01	p<0.005
Number of ions	3393	2922	2841	2692	263
Number of ions decreasing	1764	1470	1419	1300	89
Number of ions increasing	1629	1452	1422	1392	174
Expected by chance	N/A	169	67	33	16

6.4.2.3 Tentative identification of PPG from the DI-HRMS spectra

DI-HRMS spectra of sewage outfall DOM extract shows that there are clear series of evenly spaced ions with a normal distribution of ion abundances, indicative of an homologous series of compounds. When these ions were compared, the mass difference found to be separate these ions were 58 Da, which as discussed in Chapter 3 is can be indicative of PPG. However, this is only a tentative assignment until characteristic HPLC-HRMS/MS spectra. To further

investigate these series, the pattern matching algorithm described in Section 2.3.2.2 was used to find all series of ions related by this specific mass difference. Twenty-nine series of ions with a mass difference of 58.041 Da were identified which included 227 ions.

An attempt was made to calculate the molecular formula assignment all 29 oligomeric series; however, this was only possible for 10 oligomeric series of ions. Figure 6.7 shows some narrower mass windows around some of the ions for which the molecular formula could not be assigned. Figure 6.7(a) shows an example of the ions which were all found to be part of 4 of the oligomeric series. The mass difference of the consecutive ions in Figure 6.7(a) is *m/z* 0.5016, which is half the mass difference between a ^{12}C and a ^{13}C atom, suggesting that the ions from 4 of the series are doubly charged. Figure 6.7(b) shows an example of 4 ions which were found to be associated with 4 series of PPG. The mass difference between consecutive ions in Figure 6.7(b) is *m/z* 0.3344 which is a third of the mass difference between a ^{12}C and a ^{13}C atom, indicating that these ions are triply charged. Therefore, the DI-HRMS spectra were searched for ions which had a mass difference of 29.0209 Da and 19.3473 Da to find series of doubly and triply charged ions of series of compounds with a 58.0419 Da mass difference. For the ions found by searching for these mass differences the *m/z* was doubled and tripled and the molecular formula calculated to take into account the charge of the ion.

At Llanrwst 10 singly charged series of ions, 5 doubly charged series of ions and 3 triply charged series were found which could be indicative of PPG. Table 6.3 summarises the 18 series identified from the DI-HRMS spectra. PPG Sequences 1, 2, 3, 6 and 8 from Llanrwst share the same molecular formulae as the PPG series identified in Chew Stoke sewage treatment works PPG series; 3, 16, 9, 2 and 15, respectively, however, these ions span different molecular weight ranges.

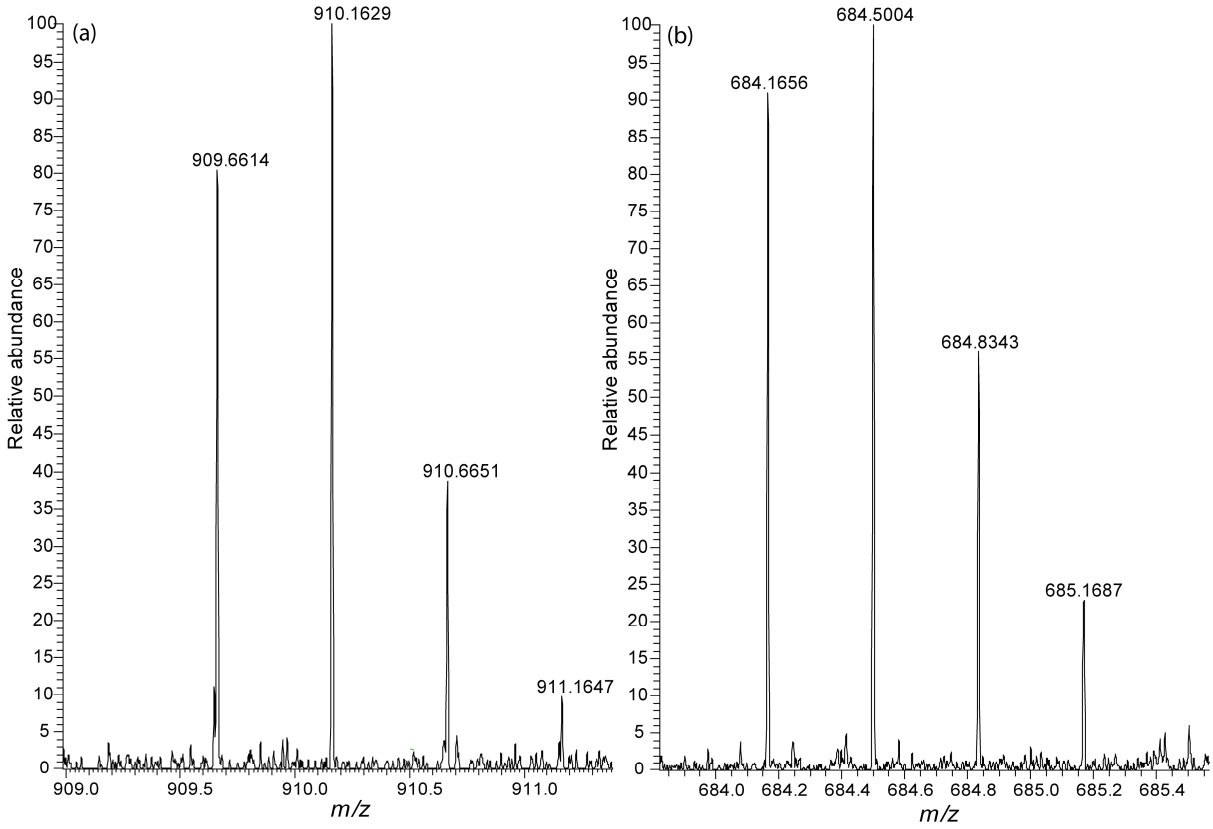


Figure 6.7. (a) DI-HRMS spectra of the sewage outfall DOM SPE extracted from Llanrwst mass range *m/z* 909.0 to 911.3 example of doubly charged ions, and (b) DI-HRMS spectra of the sewage outfall DOM SPE extracted from Llanrwst mass range *m/z* 683.8 to 685.6 example of triply charged ions.

Figure 6.8 shows the sewage outfall DI-HRMS spectra with four of the series of ions tentatively identified as PPG highlighted in red. Figure 6.7(c) and (d) displays one of the doubly and triply charged series of ions tentatively identified as PPG. All of the series tentatively identified as PPG, including the triply and doubly charged series, show a normal distribution of ion abundances typical of a synthetic polymer. However, for the series presented in Figure 6.7(a) the ion intensities have a bimodal distribution.

Of the 227 ions tentatively identified as belonging to one of the 18 series of PPG, only 23 of these have a p value < 0.005. In comparison, at Chew Stoke it was found that of the 179 ions which formed the 18 PPG, series 157 were found to have a p value < 0.005 and in Betws-y-Coed, all 3 PPG series of 24 ions were found to have a p value < 0.005. Therefore, in comparison to the previous two sewage treatment works studied, fewer of the ions which were tentatively identified as PPG were found to be significant in discriminating between the upstream and sewage outfall spectra at Llanrwst,

Table 6.3. Table of PPG sequences found at Llanrwst, outlining mass ranges of ions in each series, the range of the number theoretical monomer units, the range of P values for each series, molecular formula of end groups and adducts, possible isotopes and the charge.

PPG Series	Masses observed in spectra (*m/z*)	Masses of PPG series (Da)	Monomer units	P min value	P max value	End Group and Adduct†	Isotopes	Charge
S1	251.1855 to 1121.8151	N/A	4 to 19	0.00193	0.00912	H_3O		1
S2	368.2727 to 1064.7740	N/A	6 to 18	0.00447	0.00921	H_3O	^{13}C	1
S3	614.4834 to 1310.9841	N/A	10 to 22	0.00681	0.00919	CNH_8		1
S4	597.4568 to 1235.9171	N/A	10 to 21	0.00717	0.00921	CH_5		1
S5	175.1313 to 697.5088	N/A	3 to 12	0.00339	0.00919	H		1
S6	157.0861 to 563.3762	N/A	3 to 9	0.00193	0.02625	$NaOH_2$		1
S7	233.1748 to 671.4565	N/A	4 to 11	0.00447	0.00921	HO_2		1
S8	157.0861 to 790.5886	N/A	2 to 13	0.00193	0.00919	NOH_6		1
S9	384.2958 to 1236.9206	N/A	6 to 21	0.00717	0.00919	CH_5	^{13}C	1
S10	707.5171 to 1253.9461	N/A	12 to 21	0.00447	0.00919	CNH_8	^{13}C	1
S11	445.8284 to 1200.3710	891.6869 to 2400.7419	15 to 41	0.00525	0.00921	OH_4	^{13}C	2
S12	415.3163 to 763.5667	830.6325 to 1527.1335	14 to 26	0.00681	0.00921	CH_6		2
S13	444.8386 to 764.0686	889.6772 to 1528.1371	15 to 26	0.00604	0.00919	CH_6	^{13}C	2
S14	503.3814 to 677.5072	1006.7628 to 1355.0144	17 to 23	0.00717	0.00921	CH_6	$^{13}C_2$	2
S15	388.2876 to 1142.8294	776.5752 to 2285.6587	13 to 39	0.00525	0.00921	OH_4	$^{13}C_2$	2
S16	471.3459 to 800.2499	1414.0377 to 2400.7496	24 to 41	0.00306	0.00921	OH_5		3
S17	471.6792 to 839.2782	1415.03764 to 2517.8345	24 to 43	0.00525	0.00921	OH_5	^{13}C	3
S18	491.3618 to 820.2654	1474.0853 to 2460.7963	25 to 42	0.004472	0.00919	OH_5	$^{13}C_2$	3

† Without further structural characterisation it is not possible to differentiate between the formula of the adduct or end group using the exact mass. Therefore, the formula are grouped.

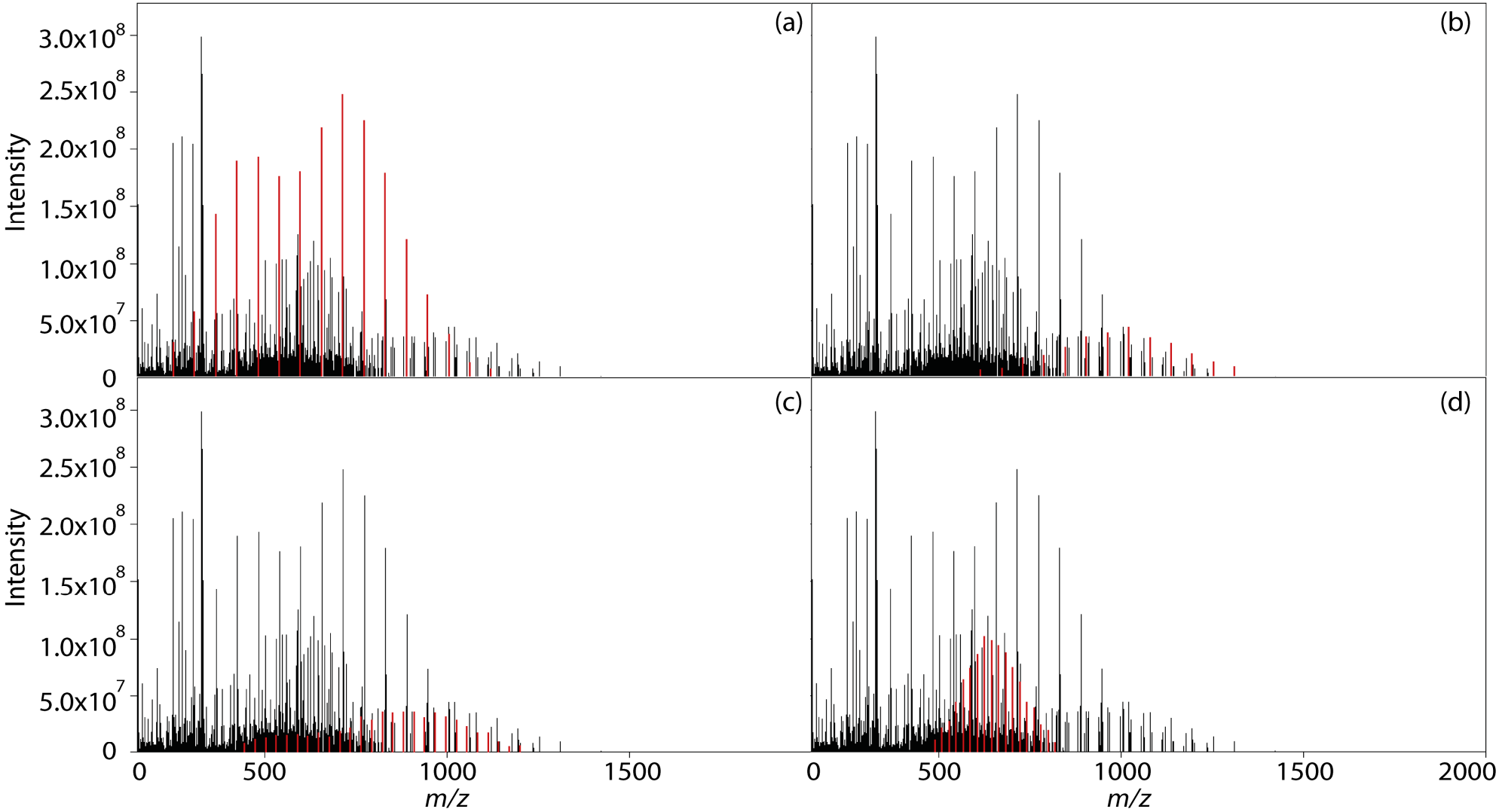


Figure 6.8. DI-HRMS spectra from Llanrwst sewage treatment works sewage effluent SPE extract with PPG sequences highlighted (red): (a) singly charged PPG series 1, (b) singly charged PPG series 3, (c) doubly charged PPG series 11, and (d) triply charged PPG series 18.

6.4.3 High performance liquid chromatography Orbitrap mass spectrometry of DOM extracted from Llanrwst

SPE extracts were analysed using HPLC-MS analyses as described in Section 2.3.3 to further investigate differences in composition and identify compounds of interest. Figure 6.9 shows the TIC of the blank, upstream, sewage outfall and downstream DOM extracted from Llanrwst. There are two early eluting peaks present in the TIC of the blank, upstream, and downstream extracts, which are not observed in the sewage outfall DOM extract eluting at 52 secs and 1 min 30 sec. There is a clear difference when comparing the TIC of the blank SPE extract and the other extracts. Furthermore, the sewage outfall TIC is clearly different to the other three extracts. However, due to the high amount of coelution it is difficult to provide a meaningful comparison of the DOM extracts using the TIC.

Figure 6.10 shows UV absorbance chromatograms for the wavelength 254 nm which has been correlated to the concentration of DOM in natural systems (De Haan and De Boer, 1987; Weishaar *et al.*, 2003; Yates *et al.*, 2016). The blank extract shows a peak eluting at 53 sec which is observed in all SPE extracts which corresponds to the peak in the TIC eluting at 52 sec. Furthermore, there is also large peak at 37 min 30 sec which corresponds to the peak seen in the TICs of the riverine and blank SPE extracts. In the blank there are no other features in the chromatogram apart from an increase in absorbance which is correlated to the increasing concentration of acetonitrile for the elution gradient. As shown in Chapter 4 and 5, there is a large unresolved peak of material which absorbs at this wavelength eluting between 2 to 18 min. This was observed across all SPE extracts of all water samples collected from all the three sewage treatment works studied.

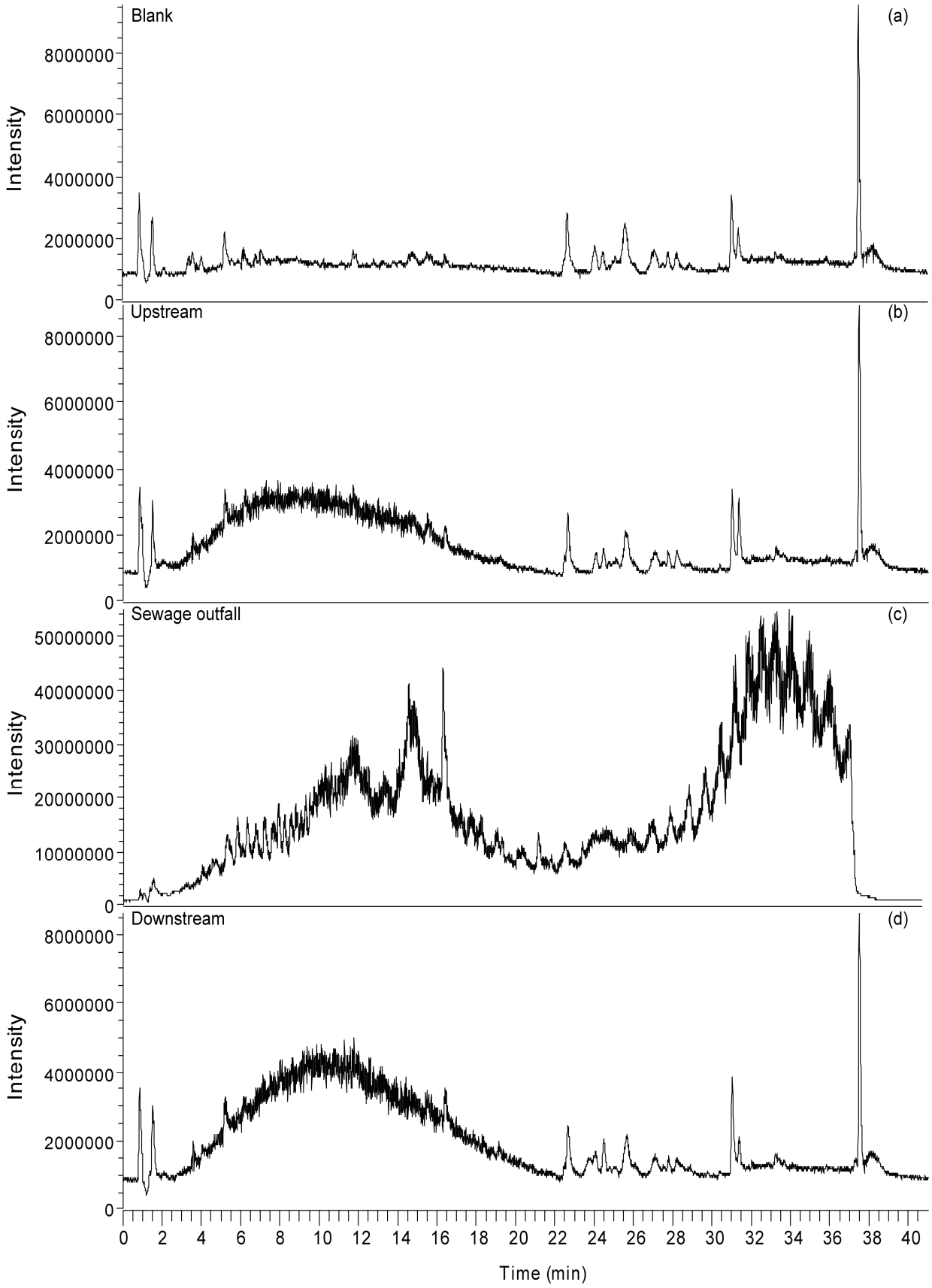


Figure 6.9. TIC from the HPLC-MS analysis for each of the SPE extracts from Llanrwst: (a) Blank, (b) upstream, (c) sewage outfall, and (d) downstream.

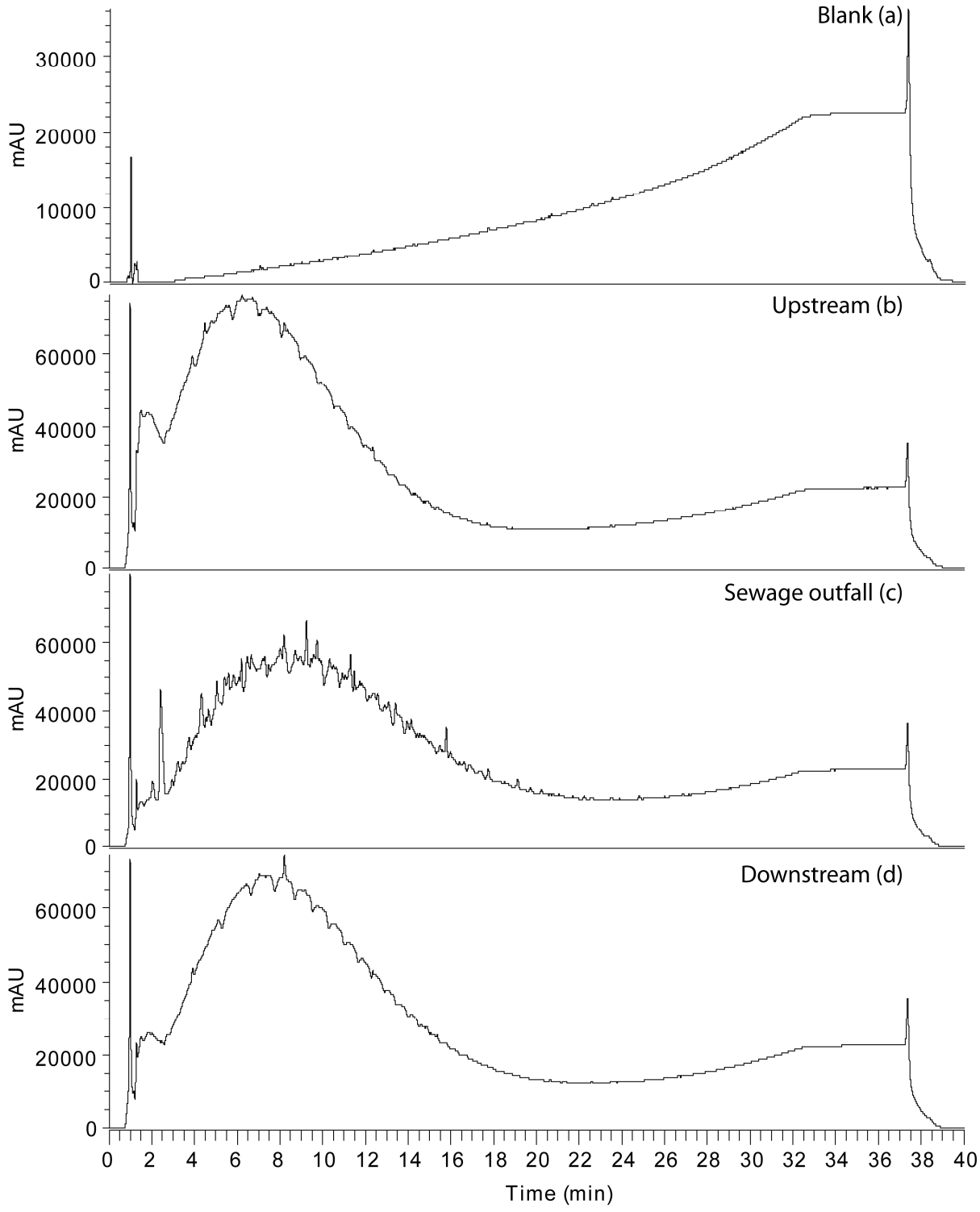


Figure 6.10. Specific UV absorbance chromatograms for the wavelength 254 nm from the HPLC-MS analysis of each of the SPE extracts from Llanrwst: (a) blank, (b) upstream, (c) sewage outfall, and (d) downstream.

Peak picking and alignment as described in Section 2.3.3.1 13942 peaks found across the blank, upstream, downstream and sewage outfall SPE extracts. All peaks found in the blank were removed from the peak table leaving 12729 peaks in total aligned across all the samples.

The ternary plot comparing the ratios of the peak areas of the EIC of each ion found in the HPLC-HRMS analyses in the SPE extracts from Llanrwst in Figure 6.11. The position of an ion on the plot allows a comparison to be made between the extracts.

11308 of the peaks detected in the HPLC-MS analysis are in the green highlighted Section of the ternary plot. This highlights peaks compounds which are within this region of the ternary plot are likely to be originating from the sewage outfall. The majority of the ions plot towards 100 % showing that these ions have a larger peak area in the sewage outfall and a lower peak area in the downstream and upstream SPE extracts. However, there are ions which plot along the sewage outfall axis. This shows that these ions have a large peak area in the sewage outfall, a smaller corresponding peak in the downstream and a smaller or no peak found in the upstream SPE extract. Ions plotting along the sewage outfall axis do not pass 50 % showing that there are no ions which have a higher peak area in the downstream in comparison sewage outfall and upstream. This was expected as the compounds input by the sewage outfall are diluted when discharged into the river. This shows that there are ions which are found in both the sewage outfall and downstream extracts and either in lower concentration or absent from the upstream extracts. This suggests that using multivariate statistics to analyse the HPLC-MS data might be able to differentiate between the upstream and downstream extracts, which was not possible using the DI-HRMS.

84 components plot along the downstream axis highlighted in blue. These ions predominantly plot in the centre of the axis showing that these are in similar abundance in both the upstream and downstream SPE extracts. No ions plot at 100 % along the axis which would indicate that these ions are unique to either the upstream or downstream, which would be unexpected. Ions plotting at the origin of the riverine axes was seen at Betws-y-Coed and the cause of this as discussed in Chapter 5 needs to be explored further.

1337 components plot in the centre in red which highlights ions which are found and have similar peak areas in all extracts. This shows that part of the molecular composition of all extracts are similar. This is likely to be because the water using as drinking water in the surrounding area and in the processing of the sewage is likely to originate from the same local water sources.

No ions plot along the upstream axis which would indicate a correlation between the upstream and sewage outfall extracts. This was not found in any of the previous studies either.

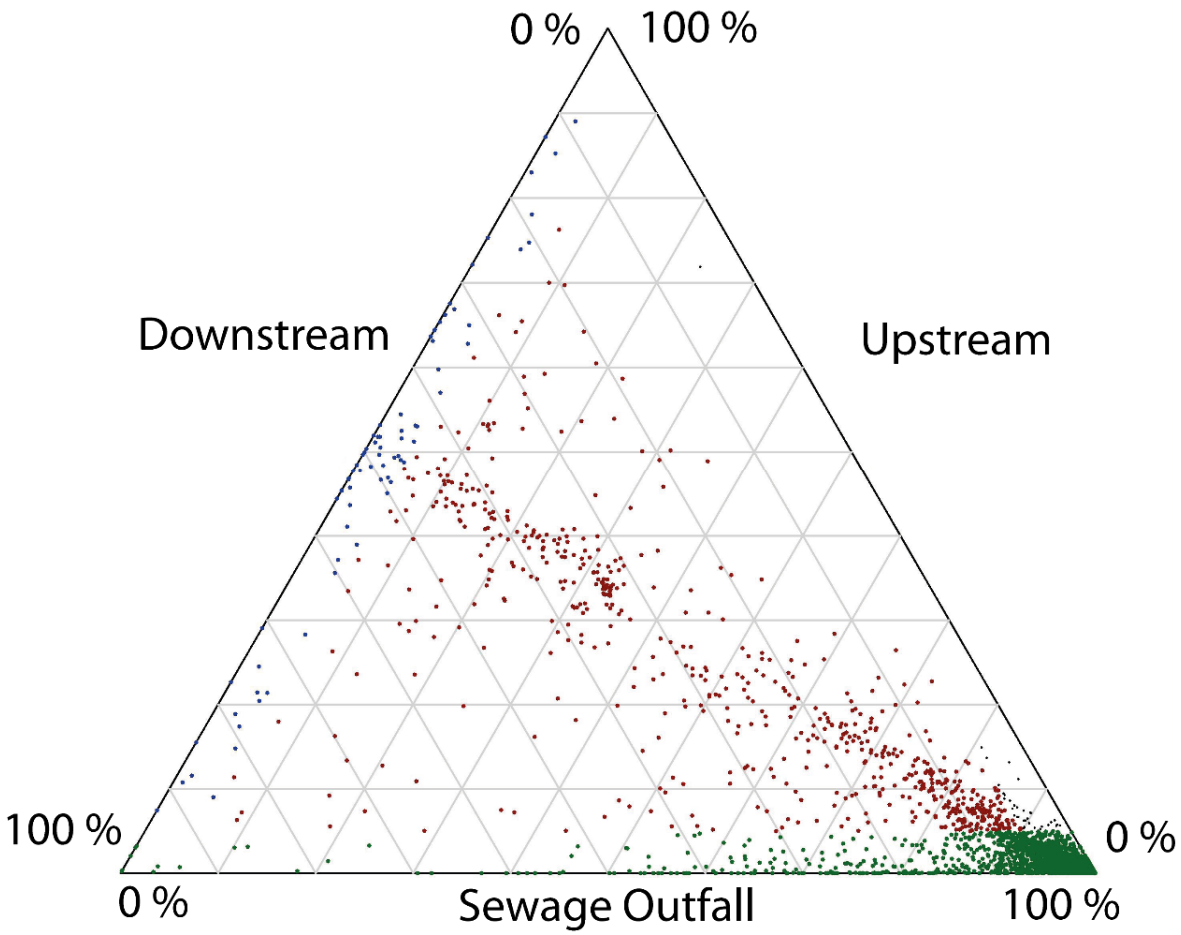


Figure 6.11. Ternary plot of ratios of the peak areas of all ions found in the HPLC-HRMS analysis, < 5 % in the sewage outfall (blue), < 5 % in the upstream (green), > 5 % in all SPE extracts (red)

As discussed in Section 3.4.3 from the Kruskal-Wallis analysis 174 ions were identified as having a p value < 0.005 and increasing intensity in the sewage outfall extracts. Masses in the HPLC-MS analysis which were within 10 ppm of the masses of the 174 ions identified as significant in the Kruskal-Wallis analysis were extracted for further analysis.

In total, 235 individual chromatographic peaks for 136 of the 174 ions were found. Figure 6.12 shows a ternary plot of all the peaks found in the HPLC-MS analysis, and the 235 which correspond to the 174 masses prioritised using in the Kruskal-Wallis analysis highlighted in red. The masses selected through the Kruskal-Wallis analysis are strongly correlated with the sewage outfall. This is reflected in the position of the mean. This shows that the Kruskal-Wallis analysis has selected peaks which have a higher peak area in the sewage outfall extracts, lower peak area in the downstream extracts and a lower/absent peak in the upstream extracts.

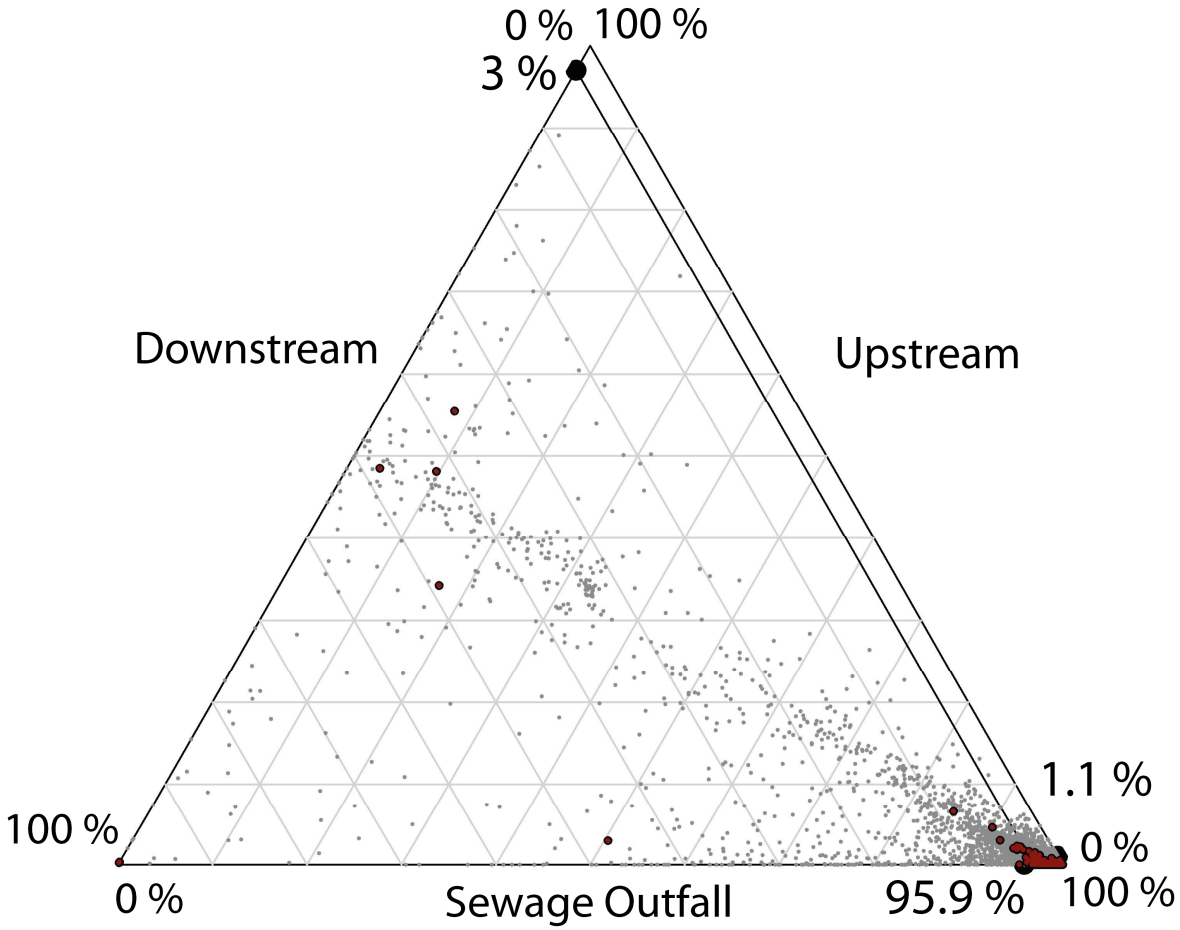


Figure 6.12. Ratios of intensities between all samples (grey) the ratios of intensities in the Kruskal-Wallis (red) and mean ratios of the upstream downstream and sewage outfalls of the Kruskal-Wallis samples

6.4.4 HPLC-MS/MS for compound identification of significant ions originating from the sewage outfall at Llanrwst

The 235 peaks were compiled into a target mass list to be analysed using MS^2 fragmentation to identify these ions. The MS^2 fragmentation was carried out as described in Section 2.3.4. In a single HPLC-MS/MS data dependant run 2028 fragmentation experiments were recorded for 338 precursor ions at different retention times. The product ions spectra of 4 ions were exact matches to database reference product ion spectra from mzCloud. These are summarised Table 6.4**Error! Reference source not found.**. The compounds identified were codeine, carbamazepine-10,11-epoxide, fexofenadine, and carbamazepine. All 4 were previously identified at Chew Stoke sewage treatment works and codeine was also identified at Betws-y-Coed.

For all the identified compounds the EICs the sewage outfall has a higher peak area indicating a higher concentration in the sewage effluent than the river. A smaller peak is seen for some of the EICs for the downstream SPE extract showing that the sewage outfall has affected the composition downstream of the point source. For all EICs there are no peaks observed for either the blank or upstream extracts. The EIC of the ion's *m/z* 300.1592 and *m/z* 253.0978 show multiple peaks for ions with the same exact mass. Therefore, these compounds may be structural isomers.

Codeine, which was found in all three sewage treatment works studied and is commonly found in the analysis of wastewater (Gros *et al.*, 2009; Santos *et al.*, 2013; Baker *et al.*, 2014). However, when comparing the EICs of the protonated molecular ion *m/z* 300.1592, for each EIC of the same exact mass, there are a number of different peaks at different retention times.

The anti-convulsant carbamazepine and its major metabolite carbamazepine-10,11-epoxide were found at Llanrwst sewage treatment works. These compounds were also identified as originating from Chew Stoke sewage treatment works (Chapter 4) and have been found in previous research studies (Miao *et al.*, 2005; Zhang *et al.*, 2008; Murray *et al.*, 2010). The antihistamine fexofenadine is a non-prescription was previously identified at Chew Stoke and discussed in Chapter 4. It has been previously identified in other sewage treatment works (Kosonen and Kronberg, 2009).

Table 6.4. Summary of the compounds identified in DOM mixture.

Precursor [*m/z*]	**Retention Time [min]**	**Fragmentation energy CID (eV)**	**P value**	**Molecular Formula**	**Compound Match**	**Database**	**Product ions [*m/z*]**
300.1592	4.11	40	0.004472	$C_{18}H_{21}NO_3$	Codeine	mzCloud	282.1489, 267.1254, 243.1033, 225.0910, 215.1067, 199.0736, 193.0648, 187.0754, 183.0804,
253.0978	7.72	30	0.003735	$C_{15}H_{12}N_2O_2$	Carbamazepine 10,11-epoxide	mzCloud	254.0817, 236.0710, 210.09187, 180.0809
502.2957	12.05	30	0.002454	$C_{32}H_{39}NO_4$	Fexofenadine	mzCloud	484.2830, 466.2726, 262.1591, 250.5923, 246.1489, 233.1174, 171.1168
237.1028	11.58	30	0.002454	$C_{15}H_{12}N_2O$	Carbamazepine	mzCloud	237.0708, 220.0758, 194.0966, 192.0810

6.4.4.1 Polypropylene glycol identification

As discussed in Section 6.4.2.3, 18 series of ions were tentatively identified as PPG from the DI-HRMS spectra of the sewage outfall. From the molecular formulae assigned to the ions, a number of isotopes, adducts and multiply charged ions were identified. The retention time can be used to identify isotopes, adducts, and the multiply charged molecular ions obtained from the HPLC-MS analysis.

The series 1, 2, 6 and 8 were identified as singly charged oligomeric series of PPG diol. The EICs of oligomers of series 1 and 2 from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst are shown in Figure 6.4. For all series there is clearly an increase in retention time as the mass of each oligomer increases. Series 1, 6 and 8 oligomers were found to coelute and the difference of mass between oligomeric ions show that these are the protonated, sodiated and ammonium adduct respectively. The PPG series 2 has a bimodal distribution of peaks (Figure 6.14 (b)). The first distribution which occurs between 10 and 27 min coelutes with the PPG diol series 1 and from the molecular formulae and difference in mass shows that these are the ^{13}C isotope series of the protonated adduct (highlighted in green). The second distribution of peaks elutes later (27 to 37 min) and all peaks have a larger peak area than the first distribution (highlighted in purple). However, the product ion spectra of the second distribution were found to contain higher molecular weight fragment ions than the precursor ion, indicating that these ions are doubly charged. The second distribution of ions in Figure 6.14 (b) couldn't be identified as PPG. The product ion spectrum of the precursor ion *m/z* 563.3762 from series 6 corresponds to the sodiated adduct of PPG diol (Figure 6.15). The product ion spectrum is similar to Figure 4.19. There are two series of product ions which correspond to the proposed fragmentation above the spectra. The product ion spectra show the characteristic 58Da spaced fragment ions. This confirms that these series are the diol of PPG.

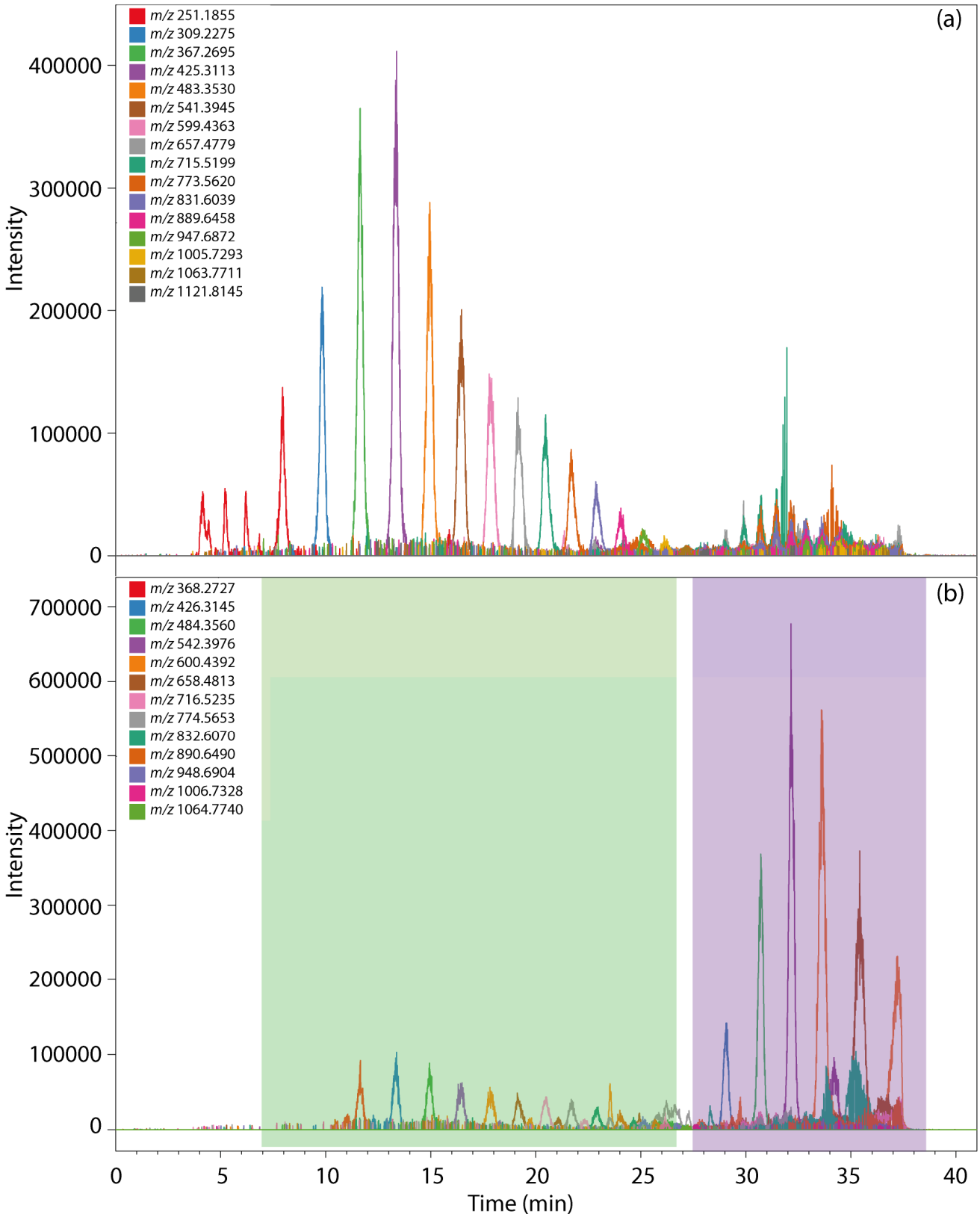


Figure 6.13. EICs from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst of the singly charged ions tentatively identified as PPG from the DI-HRMS spectra. These ions were identified as the PPG diol (a) PPG diol series 1 $[M+H]^+$, (b) PPG diol series 2 highlighting the two different distributions of peaks, earlier eluting $[M^{13}C+H]^+$ (green) and later eluting (purple)

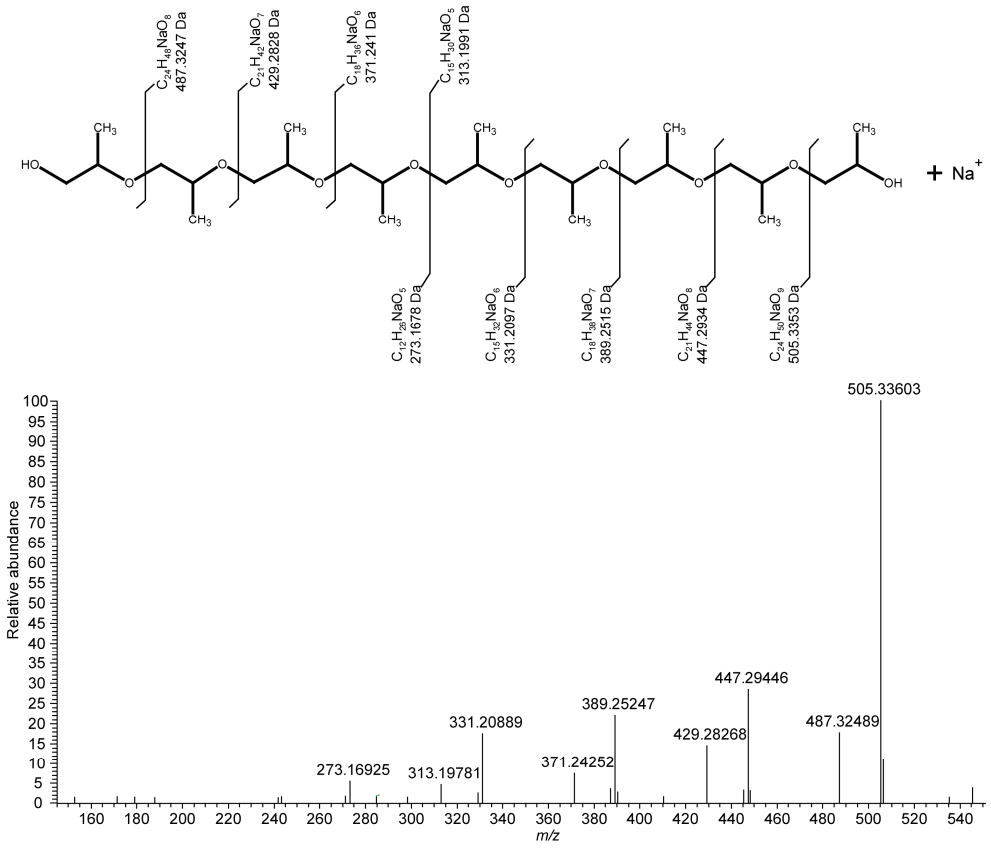


Figure 6.14. Recorded product ion spectrum of PPG diol. Precursor ion *m/z* 563.3762, retention time 16.18 min CID 30 eV of the sodiated adduct from the series S6

PPG series 11 and 15 were tentatively identified from the DI-HRMS spectra as doubly charged series of ions. From the molecular formula, the ions in series 11 is tentatively identified as the doubly charged doubly protonated molecular ions of ^{13}C PPG series 2. From the molecular formula of PPG, series 15 was tentatively identified as the doubly charged doubly protonated $^{13}C_2$ PPG diol. The EICs presented in Figure 6.15 show that the retention time of the peaks for each ion differ to the singly charged series shown in Figure 6.14 suggesting initially that these series are unrelated. However, by comparing the retention time of the peaks and calculating the molecular mass of the oligomers there is a clear continuation of the linear trend shown in Figure 6.16 (highlighted in green) of the singly and doubly charged ions. This shows that these ions are likely to be the doubly charged series of the diol oligomers. Figure 6.19(b) shows an example of a doubly charged product ion spectra of one of the ions from the doubly charged series 15. The product ion spectrum shows which the product ions have as higher *m/z* than the precursor ion confirming that the precursor ion is multiply charged. However, there are no characteristic 58 Da spaced product ions to confirm that this is PPG. Through the isolation of

the higher *m/z* product ions and by performing MS^3 fragmentation experiments, it may be possible to confirm these ions as PPG and this needs to be developed further.

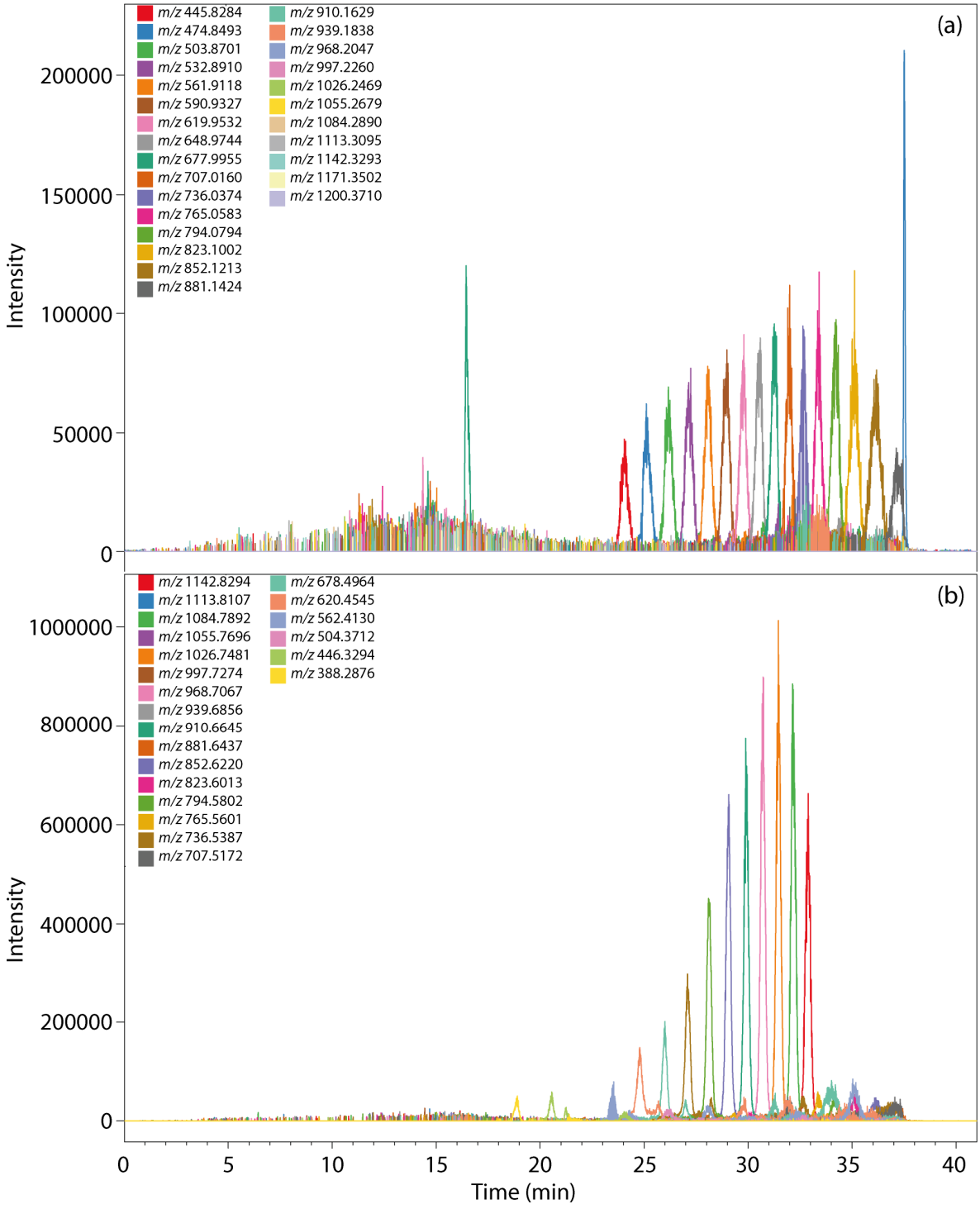


Figure 6.15. EICs from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst of the doubly charged ions tentatively identified as PPG from the DI-HRMS spectra. These ions were tentatively identified as the doubly charged PPG diol ions (a) PPG diol series 11 $[M^{13}C+H_2]^{2+}$, (b) PPG diol series 15 $[M^{13}C_2+H_2]^{2+}$.

Figure 6.16 shows that there are clearly 2 series of oligomeric peaks corresponding to series 15. One which follows the main linear trend of the diol series (highlighted in green) and the other one with a steeper slope (highlighted in purple). The two series of peaks are not clear from the EIC in Figure 6.15(b). This is due to the difference in peak height between the two series and that these peaks coelute. However, the larger peaks in Figure 6.15(b) correspond to peaks with a steeper gradient in Figure 6.16. From the EIC the larger peaks show that only every other ion is eluting when compared to the doubly charged series of ions. Figure 6.17 presents the product ion spectrum of an ion from this series which shows that there are fragment ions with a clear 58 Da spacing. By comparing this spectrum to Figure 4.17 many of the product ions are the same, showing that this is the ^{13}C, sodiated ion of PPG with a butyl and hydroxyl end group. This was also identified at Chew Stoke.

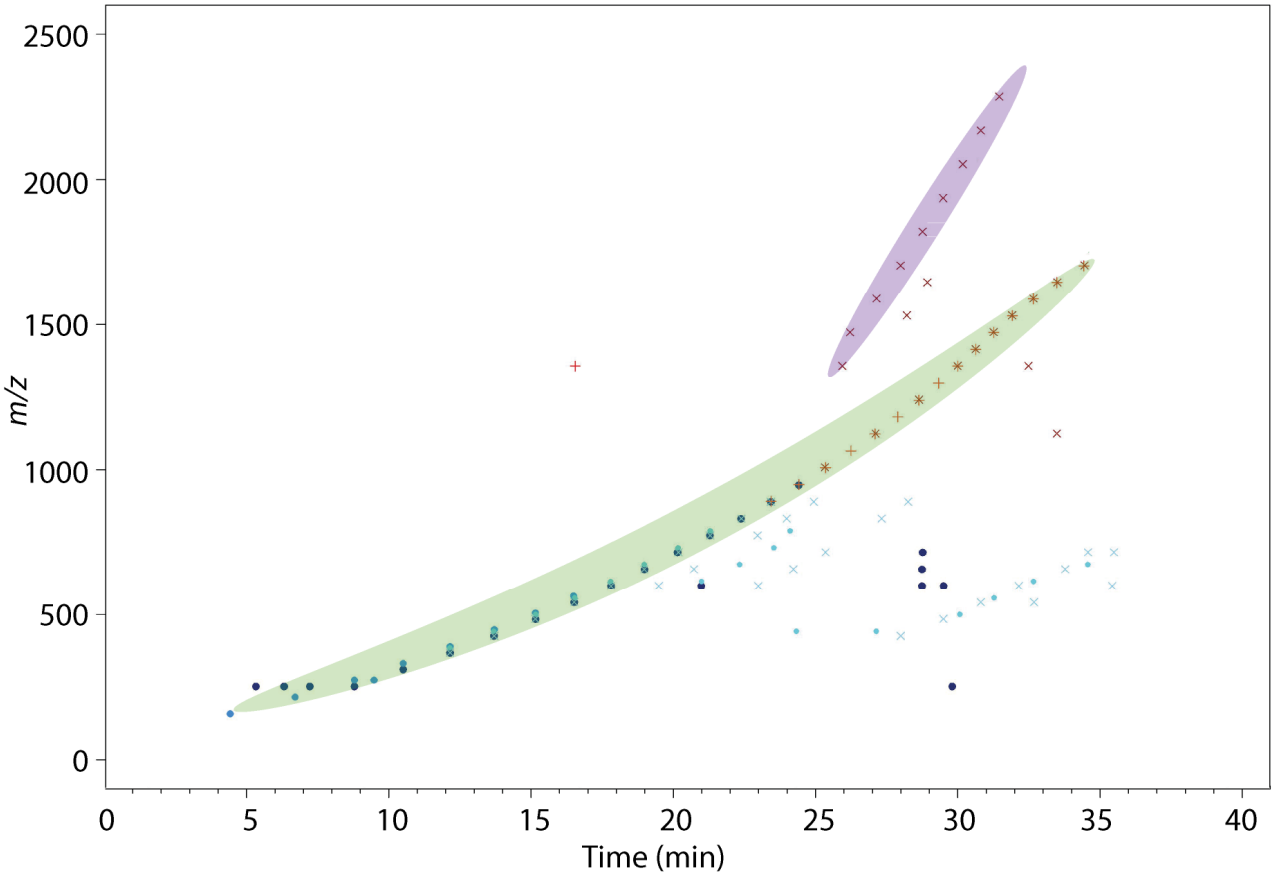


Figure 6.16. Retention time and mass of poly propylene glycol sequences which have a diol end group and different adducts, isotopes or multiple charged ions. singly charged PPG S1 (●), PPG S2 (x), PPG S6 (●), PPG S8 (●). Doubly charged ions, PPG S11(+), PPG S15 (x).

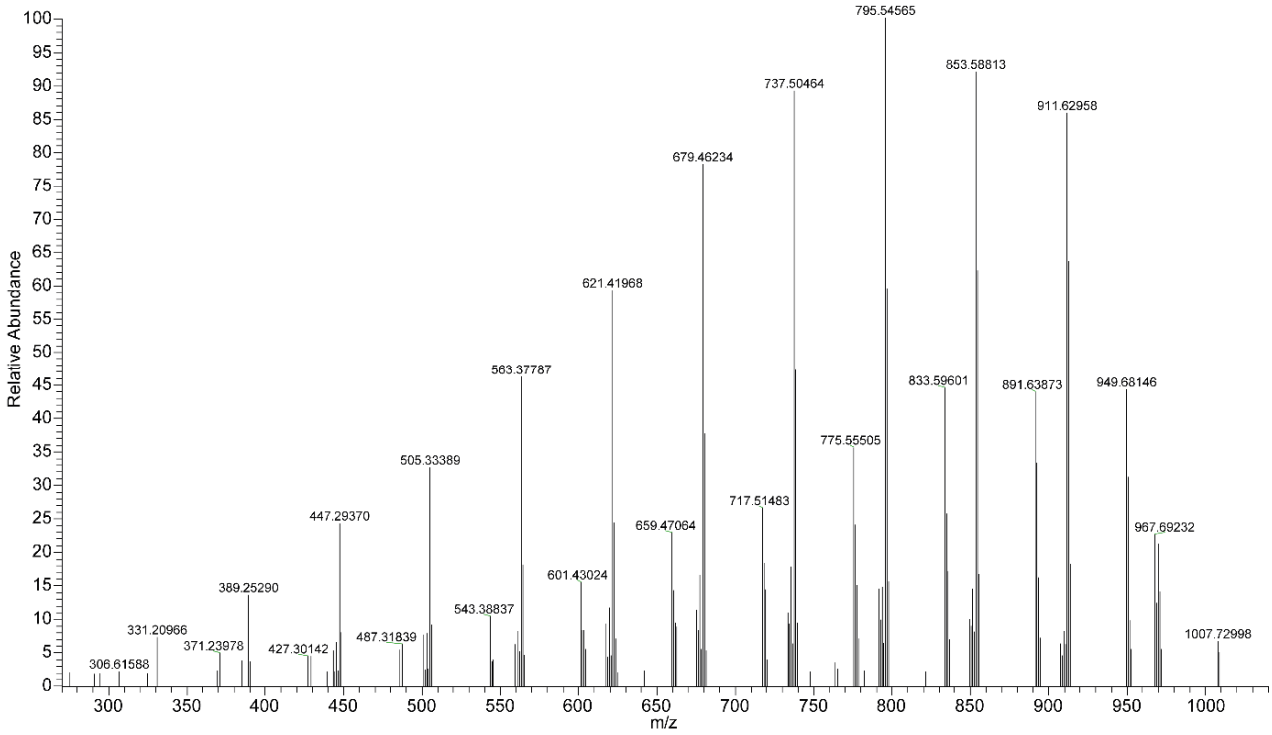


Figure 6.17. Recorded product ion spectrum of precursor ion *m/z* 1026.74 at retention time 31.56 min CID energy 60 eV.

Figure 6.18 shows an overlaid EIC of series 17, which was tentatively identified as a triply charged series of ions from the DI-HRMS spectra of the sewage outfall. The chromatogram shows that these ions have a low intensity and are poorly resolved. However, Figure 6.19 (a) presents a product ion spectrum of one of the triply charged ions. This spectrum shows that this precursor ion produces higher molecular weight product ions including the product ion *m/z* 1904.30 which is more than twice the molecular weight of the precursor ion showing that this is triply charged. No 58 Da series of product ions are produced which could be used to confirm the identity of these ions as PPG.

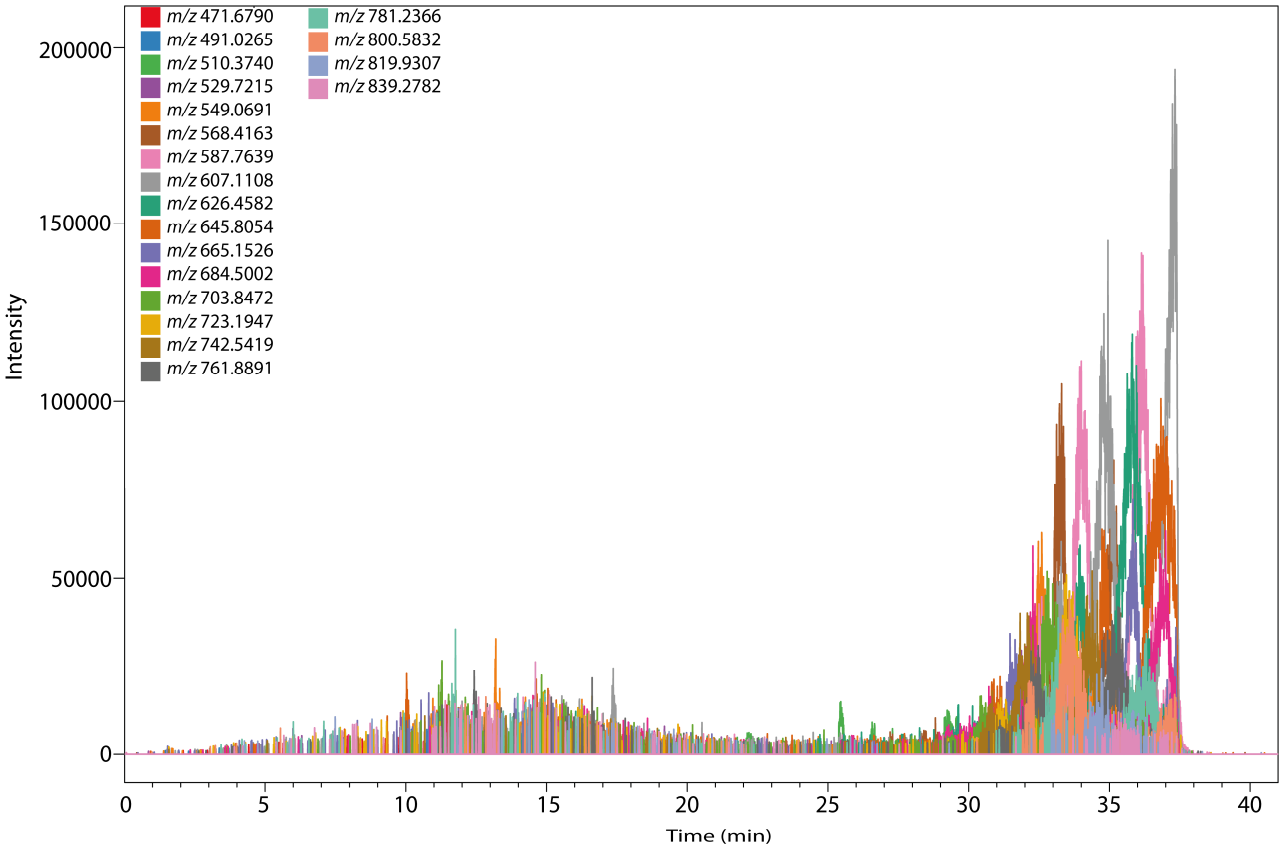


Figure 6.18. EICs from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst of the triply charged ions tentatively identified as PPG series 17 from the DI-HRMS spectra.

The PPG series 3, 4, 9, and 10 were found to elute at the same time as the series of PPG ions identified as having a butyl and hydroxyl end groups at Chew Stoke (Figure 4.15). The EIC of the oligomers of series 3 are shown in Figure 6.21. The molecular formula of 3 and 4 are the ammonium and proton adducts and series 10 and 9 are their respective ^{13}C isotope series. All of the peaks in the PPG series show a normal distribution indicative of a synthetic polymer. All of the ions in these series were found to have a p value > 0.005 therefore, no product ion spectra were recorded. Therefore, these ions are likely PPG with a butyl and hydroxyl end group however product ion spectra would need to be collected to confirm the structure

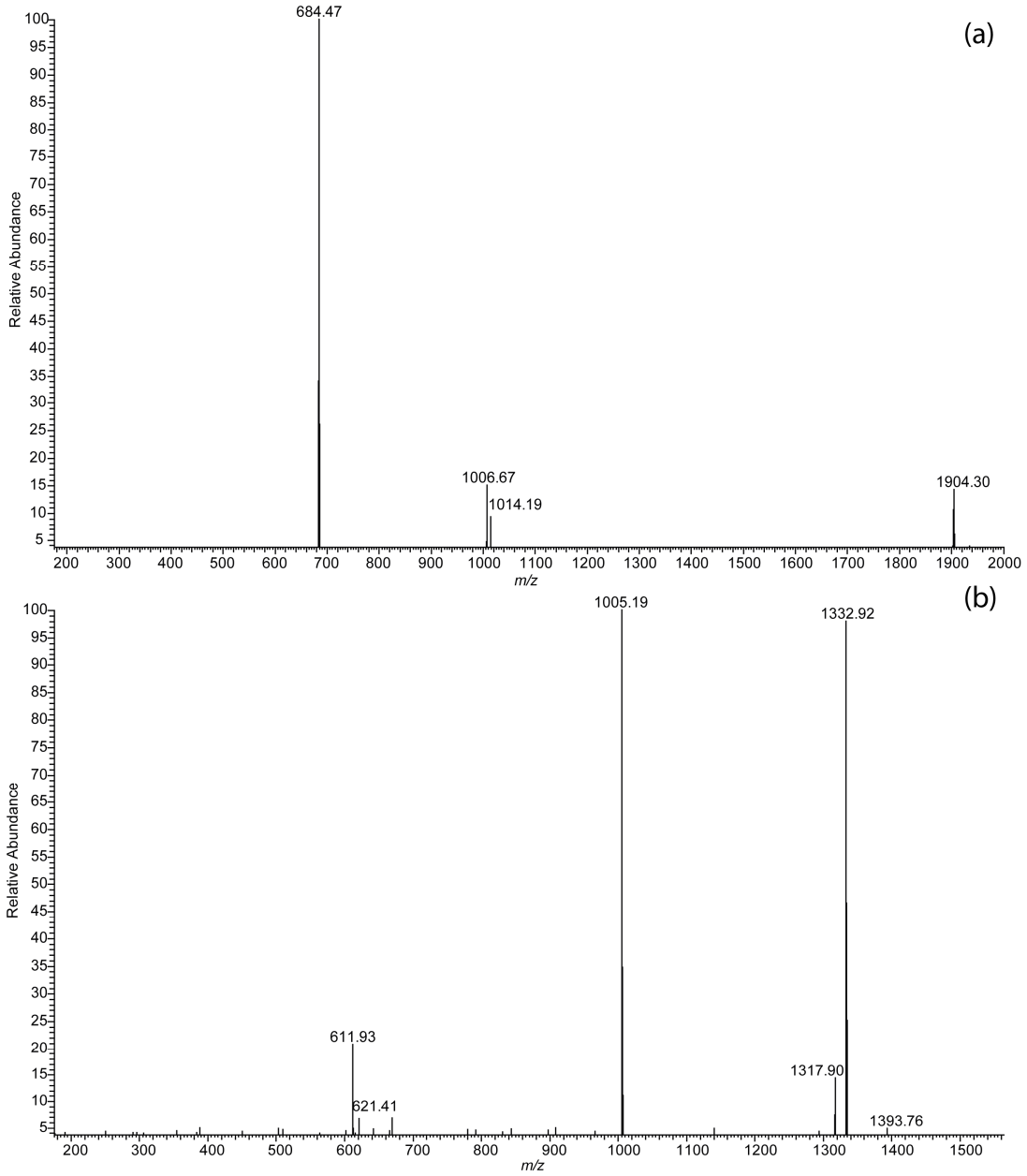


Figure 6.19. (a) Recorded product ion spectrum of the triply charged precursor ion *m/z* 684.166 at retention time 32.95 min CID energy 30 eV from the PPG series 16, and (b) Recorded product ion spectrum of the doubly charged precursor ion *m/z* 678.496 at retention time 30.92 min CID energy 50 eV from PPG series 15

.

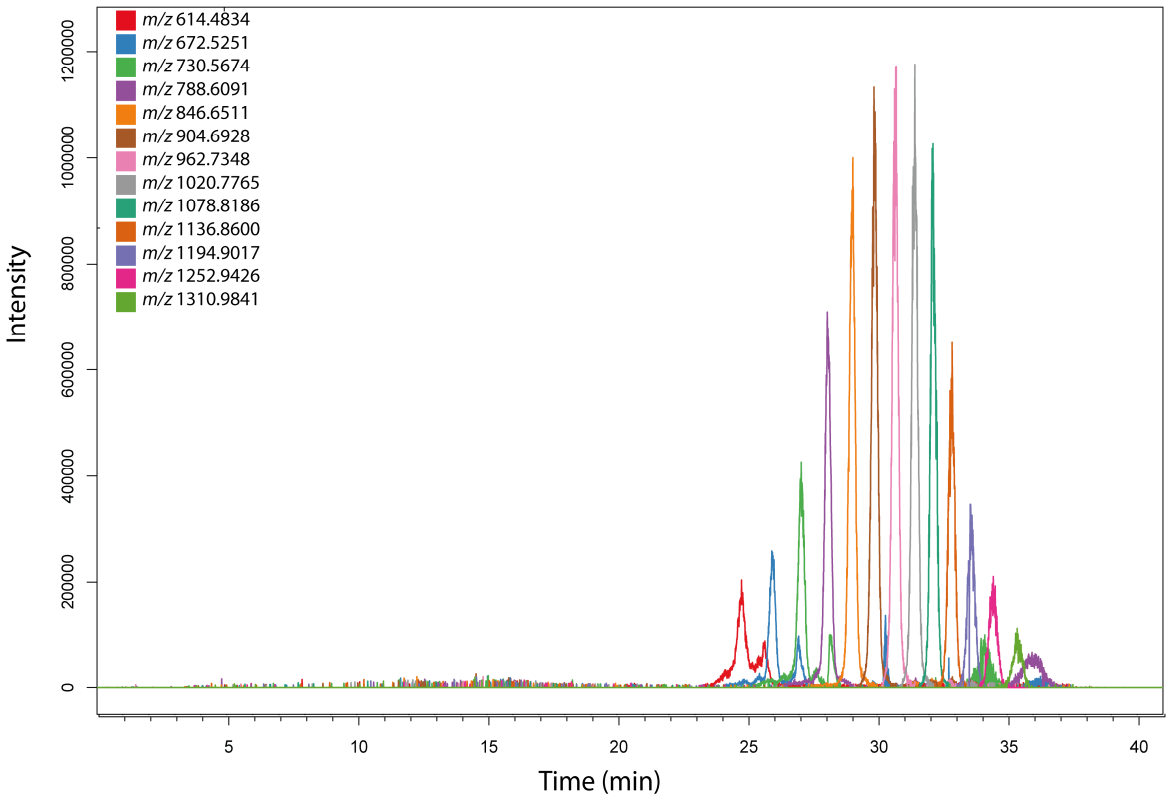


Figure 6.20. EICs from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst of the singly charged ions tentatively identified as PPG from the DI-HRMS spectra. The ions were identified as the PPG with a butyl and hydroxyl end group PPG series 3 $[M+NH_4]^+$

The PPG series 12, 13 and 14 were tentatively identified as the doubly protonated doubly charged series of PPG with a butyl and hydroxyl end group. PPG series 12 is the monoisotopic mass. Series 13 is the ^{13}C isotope of series 12, and, 14 the $^{13}C_2$ isotope of series 12 and 13. Figure 6.21 shows the EIC of the ions from series 13. These ions coelute with the ions from the singly charged series. Moreover, the mass of the doubly charged series correlate with the singly charged oligomers. Similarly, to the singly charged series no ions were found to have a p value > 0.005 therefore no product ion spectra were recorded.

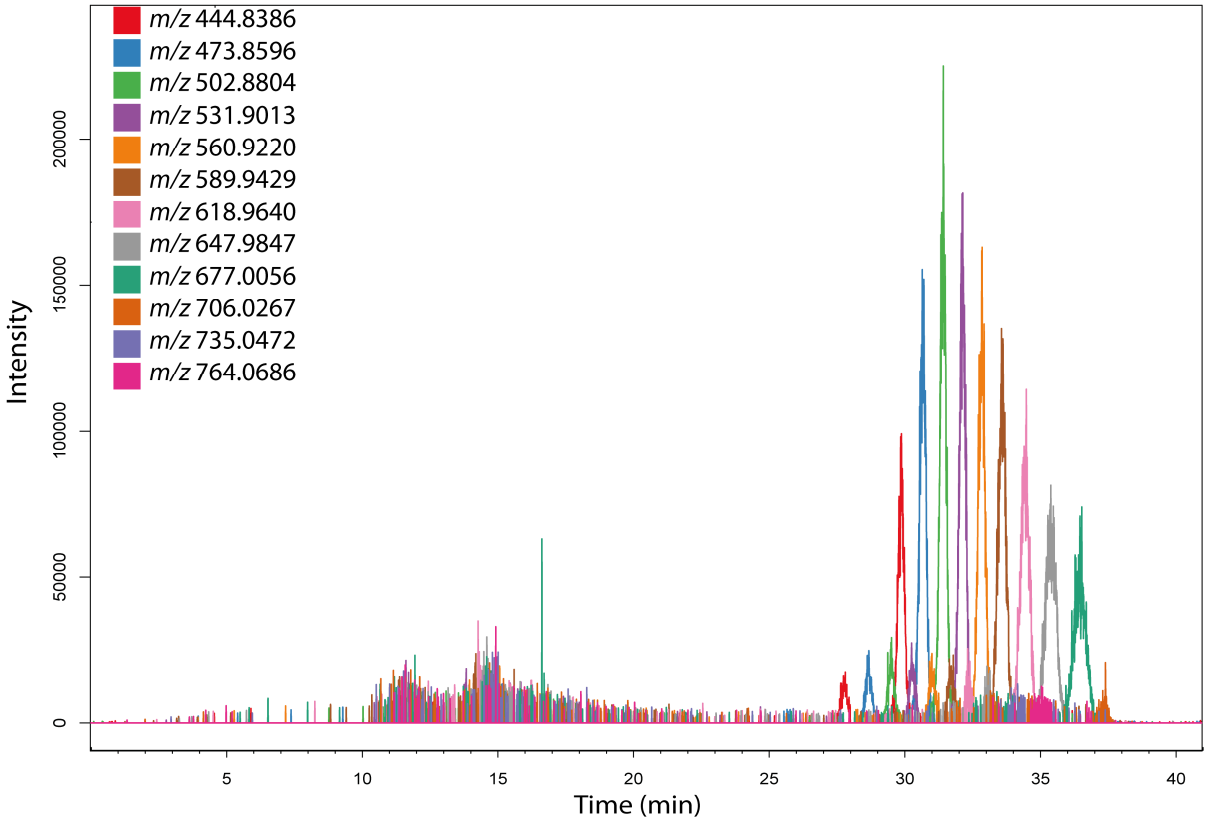


Figure 6.21. EICs from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst of the doubly charged ions tentatively identified as PPG from the DI-HRMS spectra. The ions were tentatively identified as the PPG with a butyl and hydroxyl end group PPG doubly charged with protonated adducts. The figure shows the EIC of ions from series 13 $[M^{13}C+H_2]^{2+}$

The EIC for the ions from series 7 show no normal distribution of peaks. These ions do not have a linear increase of mass with retention time. Without these characteristic features of a sythetic polymer this series of ions are unlikely to be PPG.

Figure 6.22 shows series 5 which shows two series of normally distributed peaks as expected for a synthetic polymer. Furthermore, the retention times of the ions which form this oligomeric series follow a linear increase of retention time with mass. The product ion spectra was collected for the ion *m/z* 403.229 for which there are two peaks in the EIC. One at 11.5 min (+) and the other at 12.82 min (*). Both of the precursor ions have the same exact mass however, both product ion spectra are different. The earlier eluting product ion spectra shows one major fragment ion *m/z* 331.21 but no clear series of 58 Da spaced product ions characteristic of PPG. Some of the less abundant ions have a 58 Da spacing, however, solely based on these product ions it is not possible to confirm this series as PPG. In comparison the later eluting *m/z* 403.229 ion's product ion spectra shows some fragment ions of similar abundance with a 58 Da spacing. However, from these few ions it is not possibly to unequivocally identify these spectra as PPG.

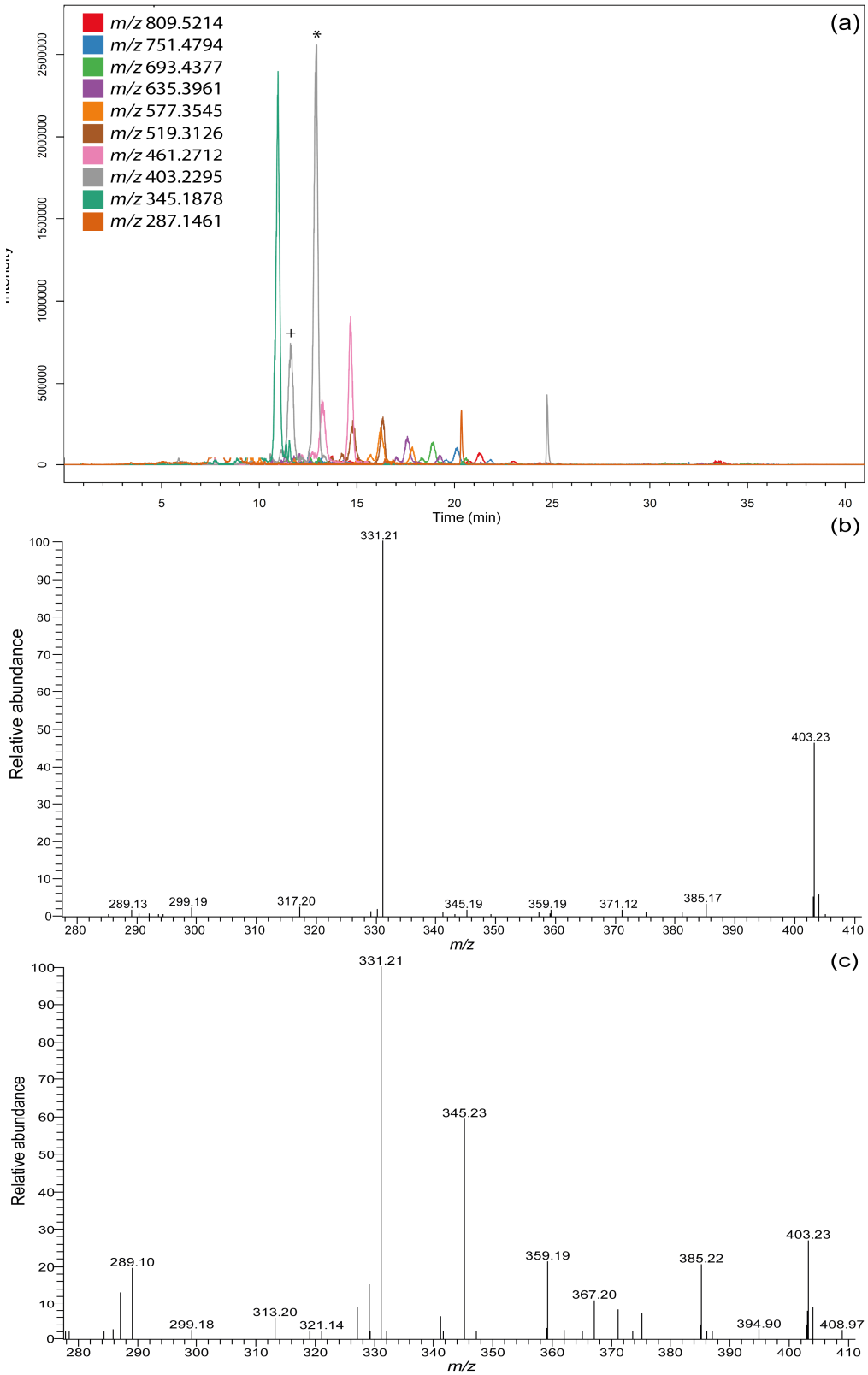


Figure 6.22. (a) EICs from the HPLC-MS analysis of the sewage outfall SPE extract from Llanrwst of the ions tentatively identified as PPG series 5, (b) Recorded product ion spectrum of precursor ion *m/z* 403.229 at retention time 11.52 min CID energy 50 eV from the PPG series 5, and (b) Recorded product ion spectrum of precursor ion *m/z* 403.229 at retention time 12.82 min CID energy 50 eV from the PPG series 5

6.5 Summary

The DI-HRMS spectra showed that there was a clear compositional difference between the extracts from river Conwy and the composition of the Llanrwst sewage treatment works outfall. Unlike Chew Stoke and Betws-y-Coed by comparing narrower mass windows of the DI-HRMS spectra it was not possible to determine if the outfall had affected the composition of the river downstream(aim i). Both the PCA and Hierarchical cluster analysis were unable to differentiate between the riverine DI-HRMS spectra indicating that the sewage outfall had not affected the composition of the River Conwy and the hypothesis stated was incorrect(aim ii & hypothesis 1). Using Kruskal-Wallis analysis a list of ions of interest which were significant in differentiating between the riverine and outfalls composition was found (aim iii).

The ternary plot of the peaks detected using HPLC-MS analysis showed that there were components which were associated with both the downstream and sewage outfall extracts which were in low abundance or absent in the upstream extract. This showed that the sewage outfall had affected the composition of the river Conwy and shows that hypothesis 1 is correct. However, only by using HPLC-MS was this difference able to be determined. By using the discriminating ions found using the Kruskal-Wallis analysis, 4 pharmaceuticals were identified and PPG diol oligomers (aim iv).

6.6 Conclusions

This chapter presents the results from using DI-HRMS and HPLC-MS to investigate the molecular composition of the effluent being discharged into the river Conwy by Llanrwst sewage treatment works. The major findings are:

(i) A direct comparison of the DI-HRMS spectra showed that the upstream and sewage outfall SPE extracts were different and the molecular “fingerprint” of the downstream was more similar to the upstream DI-HRMS spectra. (aim i)

(ii) Comparing individual ions showed that unlike Chew Stoke (Chapter 3) and Betws-y-Coed (Chapter 5) the ions found in the sewage outfall were either present in all three spectra or only in the riverine or sewage effluent extracts. Therefore, it was not possible to determine if the sewage outfall had affected the composition of the River Conwy. (aim i)

(iii) Multivariate statistical analysis using, both PCA and hierarchical cluster analysis showed that the sewage effluent extracts were significantly different from the riverine extracts. However, both analyses were unable to differentiate between the upstream and downstream DI-HRMS spectra. Heatmapping of the ions' intensities confirmed the similarity between the riverine extracts. Furthermore, this highlighted the differences between the riverine and sewage effluent DOM DI-HRMS spectra. (aim ii)

(iv) The DI-HRMS spectra of the sewage effluent extract showed there were clear series of ions with a regular mass difference. These series were tentatively identified as PPG. Using the pattern matching algorithm developed in Chapter 3, 10 series of singly charged, 5 series of doubly charged and 3 series of triply charged PPG ions were found.

(v) Using HPLC-MS analysis of the SPE extracts and comparing the ratios of the peaks detected across different extracts showed a correlation between the sewage outfall and downstream SPE extracts (Figure 6.11). This shows that possibly the sewage outfall had changed the composition of the River Conwy downstream of the point source.

(vi) Using the results of the Kruskal-Wallis analysis of the comparison of the upstream and sewage outfall DI-HRMS spectra, 174 ions were highlighted which had a p value < 0.005 and were found to have an increasing abundance in the sewage outfall DI-HRMS spectra. Four compounds were identified all of which had been identified in the previous studies at Chew Stoke and Betws-y-Coed. (aim iii and iv)

(vii) Of the 18 series of PPG both the four singly charged and two doubly charged PPG series were identified as the diol, and four singly charged and three doubly charged series were identified as the PPG with a butyl and hydroxyl end group. These were similarly identified as originating from Chew Stoke sewage treatment works. (aim iv)

Llanrwst is a complex downstream catchment environment with multiple tributaries and sewage treatment works. By using DI-HRMS the differences in the composition of DOM can be used to discriminate between the different point source and riverine DOM. Using the Kruskal Wallis analysis prioritised a relevant list of micropollutants originating from the outfall.

This shows that this untargeted methodology has the potential to be applied to point sources across different environmental settings.

Chapter 7

Overview, synthesis and future work

Chapter 7. Overview, synthesis and future work

7.1 Overview

Organic micropollutants in the aquatic environment are becoming a growing public concern because of the impact such compounds have on ecosystems of all types (Jobling *et al.*, 1998; Beketov *et al.*, 2013; Barkham, 2018). A wide range of pollutants are recognised as having damaging long-term effects on all organisms, thereby threatening the sustainability of freshwater ecosystems (Chopra *et al.*, 2011; Fong and Hoy, 2012; Li, 2013; Zenker *et al.*, 2014). These pollutants are derived from a range of point and diffuse sources. However, once discharged into the environment it can be difficult to determine which compounds are micropollutants and which are part of the natural DOM found in all aquatic systems. Therefore, the aim of this thesis was to use untargeted analysis to more comprehensively chemically characterise point sources against the background of DOM in the environment using Orbitrap HRMS. Without the characterisation of these compounds it is impossible to begin to assess their ecotoxicity and wider environmental impacts, as a basis for designing effective mitigation strategies to prevent them from entering these vulnerable water sources.

In view of the overall aims of the thesis: The DI-HRMS analyses described in Chapter 3 showed the complexity of the DOM SPE extracted from river water and sewage effluent. The analysis also highlighted the difficulties in manually comparing DI-HRMS spectra. However, it was shown that PCA and hierarchical cluster analysis could be used to determine that the sewage effluent and upstream extracts were least similar, and critically that the downstream DOM was a mixture of the two sources. A pooled QC allowed a pragmatic assessment of whether analytical variability or data processing had influenced the separations seen in the PCA. Additionally, Kruskal-Wallis analysis provided an essential data reduction step in prioritising the ions most significant in discriminating between the different DOM sources. Ions shown to increase in abundance in the sewage outfall DI-HRMS spectra were prioritised for further investigation. This methodological approach explored the complexity of the SPE extracts, provided data visualisation to assess the differences between the composition of extracts, and highlighted which ions were significant using univariate statistics (Aim i)

Chapter 4 used the results of the Kruskal-Wallis analysis from Chapter 3 to target ions found to be statistically significant and prioritise them for identification using HPLC-MS and HPLC-

MS/MS. Ternary plots of the ratios of areas of peaks in ion chromatograms detected in HPLC-MS analyses and aligned across the different extracts visualised trends in compositions. The relative areas of peaks in the HPLC-MS followed anticipated trends downstream due to the dilution of the sewage outfall DOM by the background river DOM, and negligible or absent peaks in the upstream DOM.

In all, twenty-two compounds were identified by HPLC-MS/MS, including pharmaceuticals, illicit drugs, plasticisers and metabolites. The majority had been previously identified as originating from sewage outfalls in target compound analyses, but this was the first time they had been detected in untargeted analyses. Interestingly, compounds not previously identified in targeted analyses were identified, namely; raltegravir and piperine. This DDA method combined with this statistical data reductive approach highlighted both previously identified compounds and novel compounds (aim ii)

An unexpected outcome of this research was Chapters 3 and 4 also showed the benefit of undertaking DI-HRMS prior to HPLC-MS to reveal compounds that would never have been considered as targets. The DI-HRMS spectra revealed prominent series of ions with an overall normal distribution and a 58 Da spacing, tentatively identified as PPG. From the TIC generated from the HPLC-MS analysis it was not possible to confirm that these were a series of oligomers due to substantial co-elution. However, upon MS/MS fragmentation, 58 Da spaced product ions were clearly seen characteristic of PPG. Thus, PPG formed a high proportion of the identified compounds originating from the sewage outfall, i.e. 141 components were identified as PPG in a number of series of oligomers differing mainly by their end-groups. This is the first time PPG had been identified as a significant component of sewage effluent. PPG has a wide range of uses as a surfactant, dispersant and in manufacturing (Rychłowska *et al.*, 2003; Zgoła-Grześkowiak *et al.*, 2006). Even though PPG is non-toxic, it could still impact upon the environment as a nutrient (Zgoła-Grześkowiak *et al.*, 2006; Hu *et al.*, 2008). PPG can be utilised as a substrate; therefore, this could potentially be a contributing factor to eutrophication. Furthermore, biological treatment either using active sludge or filter beds is designed to remove nutrients prior to being discharged into the environment. The presence of PPG at both Llanrwst and Chew Stoke shows that it is not being effectively removed. Both these treatment works use filter beds as their secondary treatment method this may indicate that this method does not effectively remove all bioavailable material or that the treatment process isn't fully working.

Chapters 5 and 6 used the methods developed in Chapters 3 and 4, showing that the new untargeted approach could be applied to sewage treatment works located in progressively more complex catchment environments very different to Chew Stoke. DI-HRMS of the SPE extracts showed the method to be consistently effective in distinguishing between riverine and sewage effluent DOM. The Kruskal-Wallis analysis highlighted significant compounds to prioritise for further investigation by HPLC-MS and HPLC-MS/MS, many of which had been revealed in the combined untargeted/targeted analysis studies performed at Chew Stoke.(aim iii)

The following section provides a comparative synthesis of the results from the untargeted/targeted analyses of the three sewage works investigated in this thesis. The different studies are compared further below.

7.2 Comparison of the results from Chew Stoke (Chapter 3 and 4), Betws-y-Coed (Chapter 5) and Llanrwst (Chapter 6)

This section compares the 3 different sites studied to compare the compositional differences found across all of the different sampling locations (aim iv):

7.2.1 DOC Analyses

Comparison of the DOC analyses of the filtered water collected from Chew Stoke, Betws-y-Coed and Llanrwst, shows that the concentration of DOC in the different river water samples ranged from 2.11 to 6.77 mg L^{-1}. This is a similar range and magnitude to that determined in the sewage outfalls namely: 3.19 to 7.26 mg L^{-1}. The similarity between the concentration of organic carbon in the river and sewage outfall makes it difficult to determine if the point source is inputting organic matter into the environment. Also using this measurement, it is not possible to determine if the organic matter is compositionally different.

Interestingly, the extraction efficiencies achieved for the DOM SPE method were shown to vary from 40.34 % to 80.07 %, and seemingly depended on the water being extracted. In particular, the extraction efficiencies for the river water extracts from Betws-y-Coed were found to be much higher than those achieved from other sampling locations. SPE cartridges use a particular polymer or silica phase which will selectively extract a fraction of the DOM from a water sample. The differences in extraction efficiency may be indicative of differing compositions depending on the proportion of the organic carbon which is applicable to being extracted by the cartridge. Therefore, for example, in the Afon Llugwy the proportion of

organic carbon extractable using the HLB SPE cartridge is proportionally higher when compared to the river Chew extracts. This demonstrates that the proportions of extractable organic carbon are different. However it is not possible to determine from the extraction efficiency or concentration whether the composition of what was extracted is qualitatively similar or different.

The chemical properties of which fractions of DOM are extracted by different SPE phases has been explored in previous studies from both an untargeted and targeted perspective. As discussed in the introduction different solid phases have been shown to extract different fractions of DOM from the same water sample (Raeke *et al.*, 2016; Li *et al.*, 2017). However, as discussed in the introduction the overall composition of DOM is unknown and therefore, determining which fractions of DOM are extracted is predominantly based on elemental composition and comparison using van Krevelen diagrams.The pros and cos are discussed in the introduction (Green *et al.*, 2014; Raeke *et al.*, 2016; Li *et al.*, 2017).

No extraction method will extract all the organic matter from a water sample. Therefore, using different extraction methods or different solid phases may extract a different fraction of DOM and the results will differ. However, applying a consistent sampling, extraction and analytical methodology allows the fractions to be compared between each site and different sewage treatment works using the method developed in this thesis.

7.2.2 DI-HRMS Analyses

The DI-HRMS spectra collected from Betws-y-Coed, Chew Stoke and Llanrwst all show differing “fingerprints”, reflecting the variability in both the composition of background river DOM and sewage effluent DOM across the different sites and treatment plants. Previous studies look at Chew Stoke, Betws-y-Coed and Llanrwst individually comparing them will allow further exploration of the validity of the hypotheses (iii-iv) stated in the introduction.

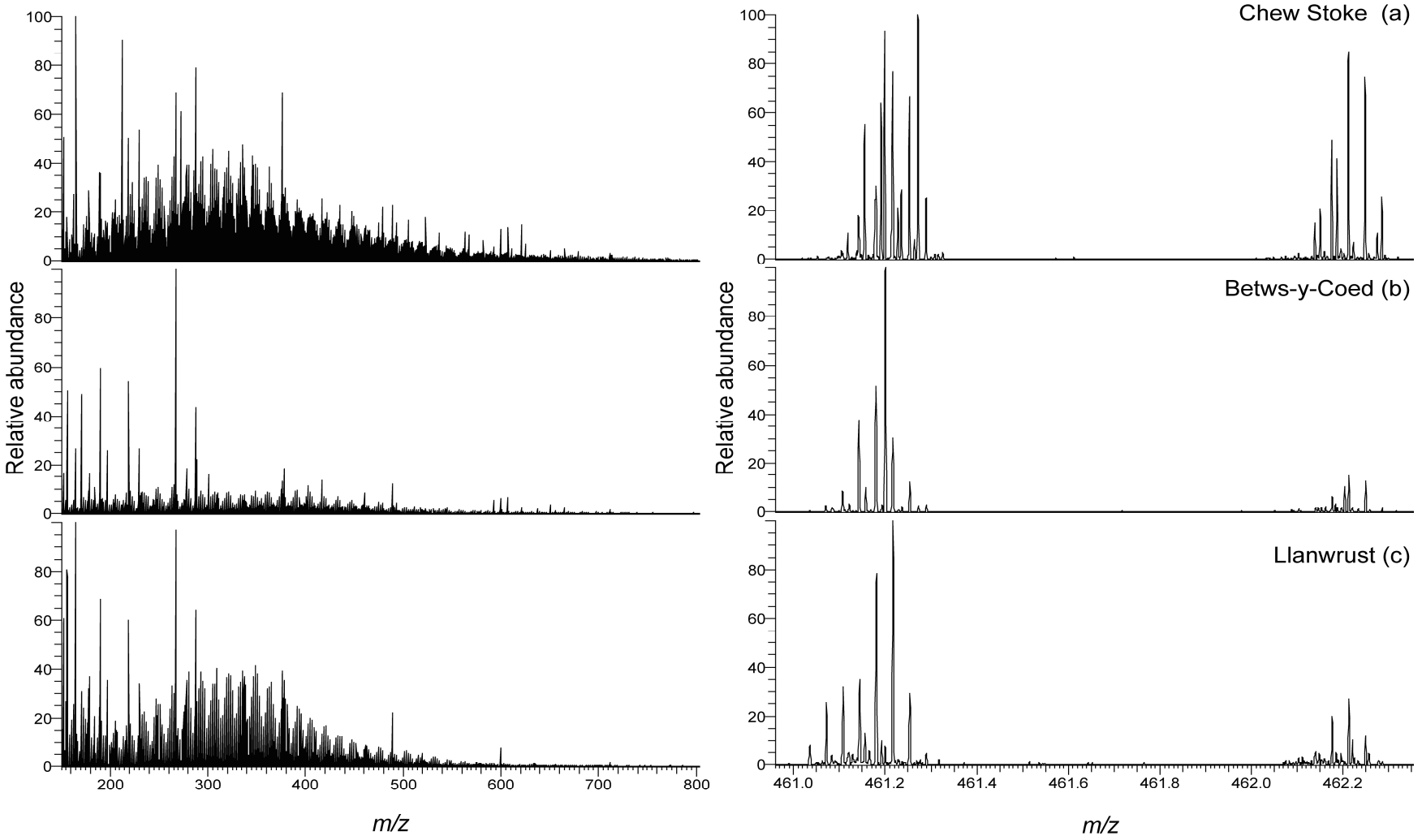


Figure 7.1. Comparison of upstream DI-HRMS spectra from (a) Chew Stoke, (b) Betws-y-Coed, and (c) Llanrwst for the mass range *m/z* 150 to 800 (left) and the mass range *m/z* 461.0 to 462.4 (right).

Comparing the DI-HRMS spectra of the upstream SPE extracts from Chew Stoke, Betws-y-Coed and Llanrwst (Figure 7.1). The upstream DOM from Chew Stoke and Llanrwst show a similar chemical "fingerprints" based on their DI-HRMS spectra, despite being from different parts of the UK. Both DI-HRMS spectra show a complex array of ions bearing similarities to published DOM DI-HRMS spectra (Flerus *et al.*, 2011; Osborne *et al.*, 2013; Mangal *et al.*, 2016). However, comparison of a narrower mass range of ions, reveals clear differences between all extracts. Both the Afon Llugwy and river Conwy contain the ions *m/z* 461.1083 and 462.1616 which are not detected at the river Chew. The river Chew contains the ions *m/z* 461.2351 and 461.2715 which are not detected in the river Conwy or Afon Llugwy. These observations show that even though initially the river Chew appears more similar overall to the river Conwy individual ions differentiate the river Chew from the rivers in the Conwy catchment. This shows the advantage of using Orbitrap HRMS to generate fully resolved DI-HRMS spectra which allow differences and similarities in the molecular composition of extracts to be readily recognised. The differences between the spectra show there is not a common molecular "fingerprint" to riverine DOM which in part disagrees with hypothesis (iv). However, some ions are detected in all three rivers therefore, part of the composition of DOM extracted from all three rivers may be the same.

Figure 7.2 compares the DI-HRMS spectra of the sewage effluent extracts collected from Chew Stoke, Betws-y-Coed and Llanrwst sewage treatment works. The chemical "fingerprints" are all unique, although Chew Stoke and Llanrwst sewage outfalls both contain prominent series of ions indicative of the presence of homologous series of compounds. However, the distributions of ions within the series are markedly different, indicating that there is unlikely to be a diagnostic "universal fingerprint" for sewage treatment works DOM. This concurs with the targeted study of Loos *et al.* (2012) which found that the absence/presence and the concentrations of detected compounds is highly variable. This disagrees with hypothesis (ii), and this variability means that the composition of each sewage treatment works maybe unique. By comparing the narrower mass ranges many of the ions are common to all three spectra. However, the ions *m/z* 279.2434 and 280.1770 are unique to Betws-y-Coed, *m/z* 280.08082 , 280.16887 are unique to Chew stoke and *m/z* 279.84016 is unique to Llanrwst. This may be a result of different treatment methods, or a fundamental difference in the composition of the sewage influent. Importantly, these DI-HRMS analyses showed the exceptional complexity of DOM, which was revealed still further by using HPLC-MS. The latter emphasises the statement made in hypothesis (i) that the compounds revealed in targeted studies represent a minor fraction of the many thousands of compounds now known to comprise aquatic DOM.

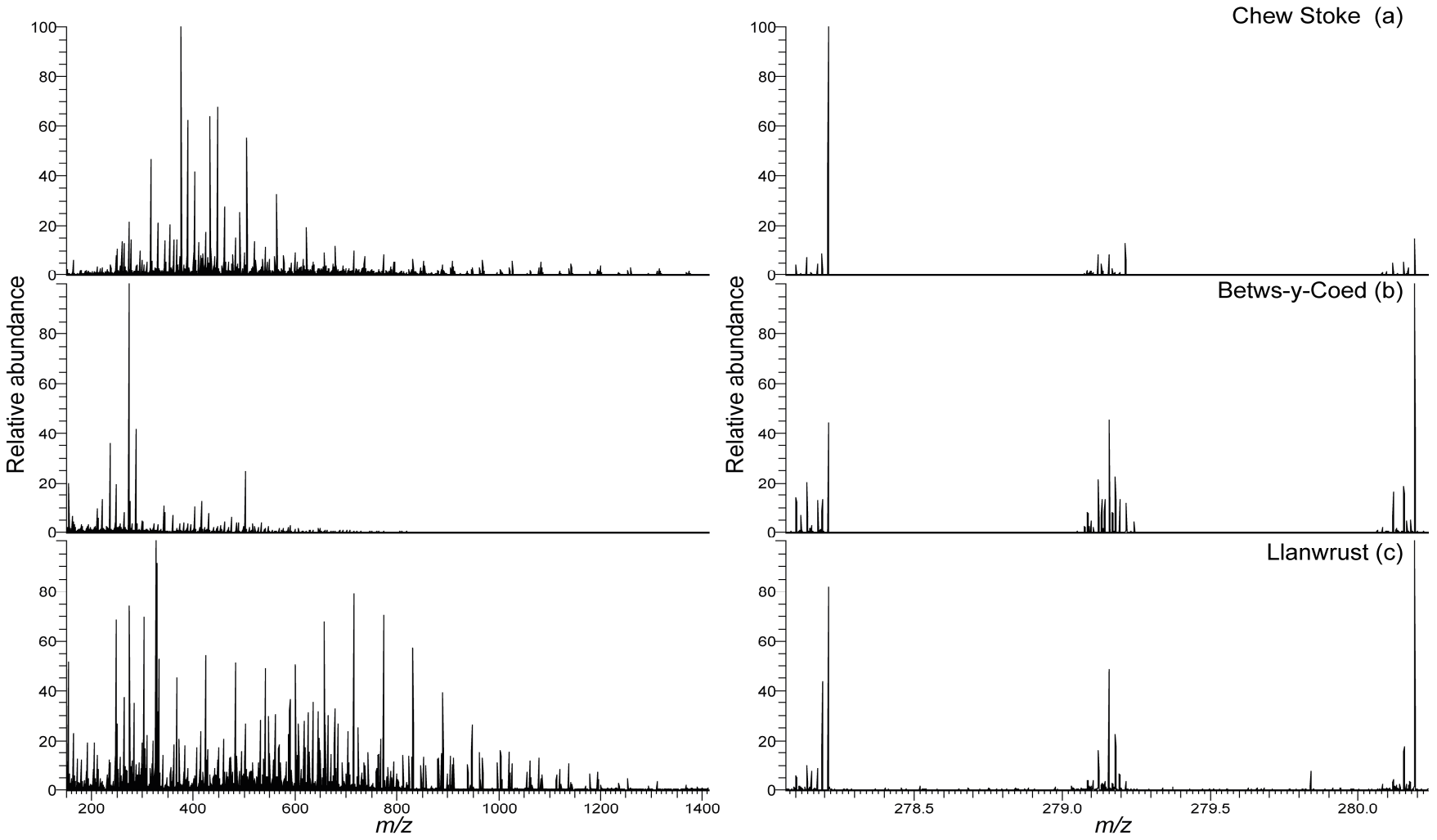


Figure 7.2. Comparison of sewage effluent DI-HRMS spectra from (a) Chew Stoke, (b) Betws-y-Coed, and (c) Llanrwst for the mass range *m/z* 150 to 1400 (left) and the mass range *m/z* 278.1 to 280.3 (right).

The DI-HRMS spectra of the downstream extracts collected from Chew Stoke, Betws-y-Coed and Llanrwst are shown in Figure 7.3. The downstream SPE extract's DI-HRMS spectra showed that Chew Stoke sewage treatment works effluent constituted significant proportion to the overall composition of the river DOM. At Betws-y-Coed the DI-HRMS spectra of the downstream DOM was very similar to the upstream DOM spectra. However, the statistical comparisons focussing on variations in ion abundances between water samples showed the contribution from the sewage outfall to be readily detectable. At Llanrwst it was not possible to definitively determine from the DI-HRMS spectra whether the sewage outfall contribution was detectable above the river DOM background due to the dilution effect of the river. By measuring the rate of discharge and the concentration of carbonaceous material from the sewage outfall in relation to the discharge rate and concentration of DOM in the receiving river this dilution effect could be calculated. A future recommendation using this method would be to determine the discharge rates of the sewage effluent and river flow. This increase in river size correlates with the dilution effect detectable in the DI-HRMS spectra: Chew Stoke, River Chew < Betws-y-Coed, Afon Llugwy < Llanrwst, River Conwy.

The multivariate statistical analyses of the DI-HRMS spectra of the riverine and effluent extracts were consistently able to show statistical differences between the sewage effluent and riverine DOM. This statistically driven untargeted methodology could be used to determine: (i) whether the DOM from a point source differs from that which is already present in a river (hypothesis v), and/or (ii) which compounds are significant in differentiating between point source and background riverine DOM. (hypothesis vi)

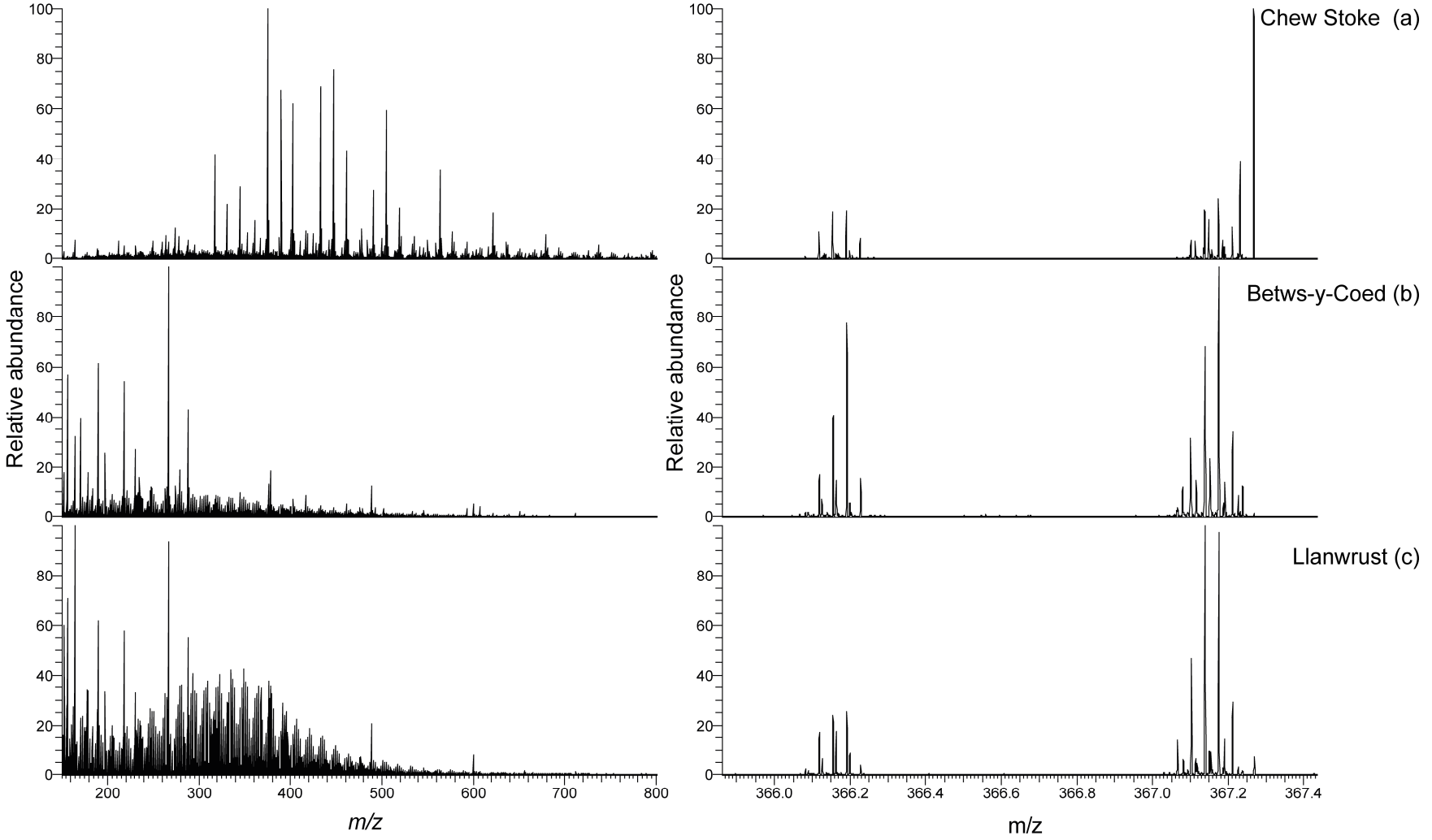


Figure 7.3. Comparison of downstream DI-HRMS spectra from (a) Chew Stoke, (b) Betws-y-Coed, and (c) Llanrwst for the mass range *m/z* 150 to 800 (left) and the mass range *m/z.* 365.8 to 367.4 (right).

7.2.3 HPLC-MS and HPLC-MS/MS Analyses

Comparing the compositions of the various DOM extracts based on HPLC-MS analysis proved impractical due to the many thousands of components detected in the DOM extracts. However, 26 compounds were identified from the sewage outfall extracts, in addition to the PPG oligomeric series. The former are summarised in Table 7.1. The compounds were found to be a mixture of compounds identified in previous studies and novel compounds as expected in hypothesis (vii-viii). This was achieved by prioritisation ions using Kruskal-Wallis analysis based upon statistically significant differences in ion abundances between the upstream riverine and sewage effluent DOM DI-HRMS spectra. Table 7.1 summarises whether an ion with the same mass as the molecular ion of one of the 26 compounds was detected in the DI-HRMS spectra of the sewage outfall, together with the p value calculated for that particular ion. The results show that many of the ions with the same exact mass as the precursor ion identified in the DOM DI-HRMS spectra were detected in the HLPC-MS analyses of the sewage effluent DOM extracts. Some of these had p values < 0.005 and would have been prioritised for identification. However, many had p values > 0.005 showing that even though such an ion may be a significant discriminator in one of the sewage treatment works studied, it was less significant in another. There are two explanations why some of the compounds which are anthropogenic origin and are significant in some sewage treatment works; (i) the compound has been discharge upstream of the outfall and the change in intensity of the ion is not large enough to make it significant or (ii) the change of intensity of the ion is not large enough to be significant. This is clearly a flaw in the method as this may exclude compounds which are biologically relevant which could be overcome by combining this method with a targeted list of compounds as well, which would expand as novel compounds are identified. Finally, some ions were detected in one but not in another sewage treatment works, which shows the variability of composition of DOM from sewage outfalls.

The EICs of the 26 ions were plotted using *m/z* of the ions presented in Table 7.1 that compares the sewage effluent extracts from Chew Stoke, Betws-y-Coed and Llanrwst. Peaks in the EIC for these ions were compared to determine if the compound identified in one of the extracts were present in the other sewage outfalls. As shown by Table 7.1, many peaks were found in other extracts with the same retention time and *m/z* but were not identified because: (i) these compounds weren't found to be statistically significant, (ii) A characteristic product ion spectra couldn't be obtained because of other ions coeluting within the isolation width of the Orbitrap,

or, (iii) the concentration of the compound was too low. However, using the retention time and *m/z* it is possible to tentatively identify these compounds as present in the other sewage effluent extracts. However, standard addition or further product ion spectra would need to be obtained to confirm the presence of these compounds.

PPG was identified at both Llanrwst and Chew Stoke sewage treatment works. The series of oligomeric PPG ions were a prominent feature of both DI-HRMS spectra of the sewage effluent extracts. In both sites PPG diol and PPG with a butyl and hydroxyl end group were identified. However, these PPG series spanned different molecular weight ranges. At Chew Stoke and Llanrwst additional series of ions with a 58 Da spacing were observed and tentatively identified as PPG. Many of these ions also produced product ion spectra which contained fragment ions spaced by 58 Da, however, from the limited number of product ions it was not possible to determine the structure of these PPG compounds. PPG was a significant contributor to the compounds identified from the DOM at both Llanrwst and Chew Stoke. The identification of the polymer shows the benefit of using both DI-HRMS and HPLC-HRMS as the oligomeric series were identifiable from the DI-HRMS where as the TIC of the HPLC-HRMS showed no distribution of peaks indicative of a synthetic polymer; only using the EIC of the oligomeric ions were these characteristic distributions able to be resolved.

PPG is claimed to be biodegradable (Zgoła-Grześkowiak *et al.*, 2006; Hu *et al.*, 2008). However, the results presented here show PPG persists through the secondary treatment methods of Llanrwst and Chew Stoke treatment works which both use trickle filter beds. In contrast, the Betws-y-Coed sewage works uses active sludge as its secondary treatment method and the PPG series identified at Chew Stoke and Llanrwst were undetectable in the Betws-y-Coed sewage effluent. There are two possible explanations which are; (i) The different treatment methods were having different effects on sewage effluent DOM composition or (ii) PPG is not present in the sewage influent which is treated by the Betws-y-Coed treatment works.

Table 7.1. Comparison of identified compounds across all the different sewage outfall DOM SPE extracts from the three different sites.

Name	Mass (*m/z*)	DI-HRMS p value Chew Stoke[†]	DI-HRMS p value in Betws-y-Coed[†]	DI-HRMS p value Llanrwst[†]	Peak detected HPLC-MS Chew Stoke	Peak detected HPLC-MS Betws-y-Coed	Peak detected HPLC-MS Llanrwst
N,N-Diethyl-3-methylbenzamide, (DEET)	192.1388	0.0014	**0.00193**	**0.003735**	Yes	Yes	Yes
N,N'-Diphenyl guanidine	212.1183	**N.D.**[*]	0.0019	**0.227638**	Yes	Yes	Yes
Carbamazepine	237.1023	0.0013	0.0019	0.0025	Yes	Yes	Yes
Carbamazepine 10,11-epoxide	253.0978	0.0013	**0.0045**	0.0037	Yes	Yes	Yes
Lamotrigine	256.0152	0.0013	**0.00193**	**0.004097**	Yes	Yes	Yes
Propranolol	260.1647	0.0013	**0.0045**	**0.133989**	Yes	Yes	Yes
Nortriptyline	264.1752	0.0013	**N.D.**	**N.D.**	Yes	No	No
Mirtazapine	266.1657	0.0013	**N.D.**	**0.005433**	Yes	No	No
Atenolol acid	268.1544	0.0016	**N.D.**	**0.005457**	Yes	Yes	Yes
Galaxolidone	273.1855	0.0013	**0.00193**	**0.0133**	Yes	Yes	Yes
Amitriptyline	278.1909	0.0013	**N.D.**	**0.006806**	Yes	No	No
Venlafaxine	278.2113	0.0013	**N.D.**	**0.06806**	Yes	Yes	Yes
Piperine	286.1443	0.0013	**0.003735**	**0.049292**	Yes	No	No

Benzoylecgonine	290.1392	0.0013	**0.003058**	**0.007521**	Yes	No	Yes
Codeine	300.1592	0.0013	0.0019	0.0044	Yes	Yes	Yes
Dihydrocodeine	302.1754	**0.0063**	0.0019	**0.008652**	Yes	Yes	Yes
Cocaine	304.1549	0.0013	**0.003058**	**0.024234**	Yes	No	Yes
Citalopram	325.1711	0.0024	**N.D.**	**0.006806**	Yes	Yes	Yes
Quinine	325.1915	0.0013	**N.D.**	**N.D.**	Yes	No	No
Desacetyl diltiazem	373.1586	0.0013	**N.D.**	**N.D.**	Yes	No	No
Cetirizine	389.1632	**N.D.**	0.0031	**N.D.**	Yes	Yes	No
Tri(butoxyethyl) phosphate	399.2512	0.0013	**N.D.**	**0.006806**	Yes	No	No
Flecainide	415.1456	0.0013	**0.005248**	**N.D.**	Yes	Yes	No
Raltegravir	445.1636	0.0013	**N.D.**	**N.D.**	Yes	No	No
Fexofenadine	502.2957	0.0013	**0.00193**	0.0025	Yes	Yes	Yes

† Bold text symbolises that the ion was not identified in this sewage treatment works. N.D. indicates This ion was not detected in the DI-HRMS spectra of the sewage outfall.

7.3 The methodology in the context of the literature

DI-HRMS and HPLC-MS applied in these three studies have shown the overall complexity of sewage effluent and DOM. The most intensive targeted studies of sewage effluent and surface water performed to date have used lists of anthropogenic compounds comprising up to only 300 analytes (Loos *et al.*, 2012; Ruff *et al.*, 2015). However, the research performed in this thesis shows how these compounds represent only a minute fraction of the total DOM inputted by a typical sewage outfall. As discussed in Chapter 1 the presence/absence of a compound and its concentration was shown to vary across different water samples. As discussed in Chapter 1 there are two main approaches used to analyse DOM using MS in the literature; untargeted and targeted analysis.

Untargeted analysis is predominantly carried out using DI-HRMS and tries to determine if there are compositional differences between extracts from different environmental settings or extraction methods. The extracts are then compared either using the calculated molecular formula of ions observed in the DI-HRMS or multivariate statistics (Osborne *et al.*, 2013; Kellerman *et al.*, 2014; Minor *et al.*, 2014; D'Andrilli *et al.*, 2015; Dubinenkov *et al.*, 2015; Li and Minor, 2015; Mosher *et al.*, 2015; Li *et al.*, 2017). However, these methods have not been applied to anthropogenic sources in the environment. Furthermore, these studies rarely aim to identify compounds which discriminate between extracts. However, the method developed as part of this thesis has demonstrated that:

(i) DI-HRMS and multivariate statistics can be used to characterise point sources in the environment.

(ii) Pooled QCs provide an effective validation method to confirm that the differences between extracts are due to differences in composition, and not as a result of pre-processing or analytical variability.

(iii) Univariate statistics can determine which ions are most significant in distinguishing between the different extracts of DOM and prioritise them for identification. This provides an essential data reductive step.

Previous studies commonly use van Krevelen diagrams to visualise the differences between different DI-HRMS spectra of DOM (Dubinenkov *et al.*, 2015; Li and Minor, 2015; Mangal *et al.*, 2016). However, this thesis shows the benefit of using multivariate statistics in particular

PCA which provides an effective method to determine if there is a difference in composition. Furthermore, the use of hierarchical cluster analysis and heatmapping visualises the changes in intensity lost when using van Krevelen diagrams.

In contrast, targeted analysis focuses on a predetermined list of compounds predicted/known to be found in surface water, aiming to confirm identifications and determine concentrations. However, the limitations of this approach is recognised, as is the requirement for improved prioritisation/data reductive methods to determine which compounds to focus upon identifying (Hug *et al.*, 2014; Ruff *et al.*, 2015). This thesis presents such a method that improves upon current targeted methods by:

(i) Prioritising compounds using Kruskal-Wallis analysis to determine which ions are significant in discriminating between extracts.

(ii) The significance of all ions is calculated when comparing extracts using Kruskal-Wallis analysis. This provides a measure of significance of a target compound in differentiating between extracts and contextualising the finding when compared to the overall composition of the extracts.

The use of Ternary plots provided an effective way of visualising the different peaks detected in the HPLC-MS analyses by comparing the ratios of the peak areas between the samples. This allowed trends to be clearly shown between the composition of the extracts. To further this work, it would be beneficial to analyse all extracts using HPLC-MS and use multivariate statistics to compare the differences between HPLC-MS analyses. Using standards of identified compounds will allow quantification either using standard addition or an isotopic standard.

7.4 Potential applications of this new method

There is still much research that needs to be carried out to improve the understanding of DOM at the molecular level. Due to the complexity, diverse number of sources and dynamicity complete characterisation of DOM is impossible. However, the methodology developed in this thesis has a wide applicability to studying different aspects of DOM in the environment.

7.4.1 Passive sampling

One of the limitations of grab sampling is that the water sample collected only represents a specific time. However, passive sampling provides an alternative method of sampling to include temporal changes in the concentration of compounds in DOM. Passive samplers use a

solid phase housed between two filters which is deployed in the aquatic system being investigated (Stuer-Lauridsen, 2005; Vrana *et al.*, 2005). As water passes through the filters and solid phase, the organic analytes in the aquatic system accumulate over time. The accumulation of a compound onto a passive sampler is split into two regimes; the kinetic and equilibrium as shown in Figure 7.4. In the kinetic regime analytes accumulate at a rate proportional to the concentration of the compound in the system. The rate of accumulation of a compound can be explained by Equation 7.1; where C_s is the concentration of the analyte on the phase, C_w is the concentration of the analyte in the water, k_1 is the rate of absorption, k_2 is the rate of desorption and t is the time of the deployment (Vrana *et al.*, 2005; Zabiegała *et al.*, 2010; Harman *et al.*, 2012). Therefore, by measuring at multiple time points within the kinetic regime the average concentration of compounds can be established over a predetermined time. The sorbent phase has a finite capacity therefore when the phase becomes saturated it is known as the equilibrium regime.

$$C_s(t) = C_w \frac{k_1}{k_2} (1 - e^{-k_2 t}) \quad \textbf{(7.1)}$$

The benefit of using passive samplers is that episodic pollution events are captured and therefore a more accurate estimation of the concentration of a compound can be gained over time. Also, compounds which are below the limit of detection in a grab sample can be concentrated over time onto the solid phase allowing their identification and quantification (Vrana *et al.*, 2005; Harman *et al.*, 2011). As the filtration and extraction of the water sample is essentially performed *in situ* this provides a fast sample preparation method for DI-HRMS or HPLC-MS analysis. The disadvantages of passive sampling are that these experiments take much longer to deliver results and the averaging of a pollutant over time means that the magnitude and duration of episodic pollution is lost. Furthermore, it has been shown that the accumulation rate can fluctuate with pH and temperature (Harman *et al.*, 2012; Jeong *et al.*, 2017)

Passive samplers have been used for targeted analysis studies to determine the concentration of selected compounds within a water system including pharmaceuticals and pesticides (Harman *et al.*, 2011; Rujiralai *et al.*, 2011; Morin *et al.*, 2013; Metcalfe *et al.*, 2014). However,

as shown by this thesis, the sorbents used for SPE can extract many different compounds. Therefore, passive samplers could be used for untargeted analysis to determine compounds derived from a point source. Furthermore, targeted studies have shown that compounds which have been pre-selected under laboratory conditions can show signs of desorption and non-linear uptake both in the laboratory and in the environment (Rujiralai, 2007). Therefore, it would be critical to understand whether the rate of uptake onto a phase is linear when deployed in an environmental setting.

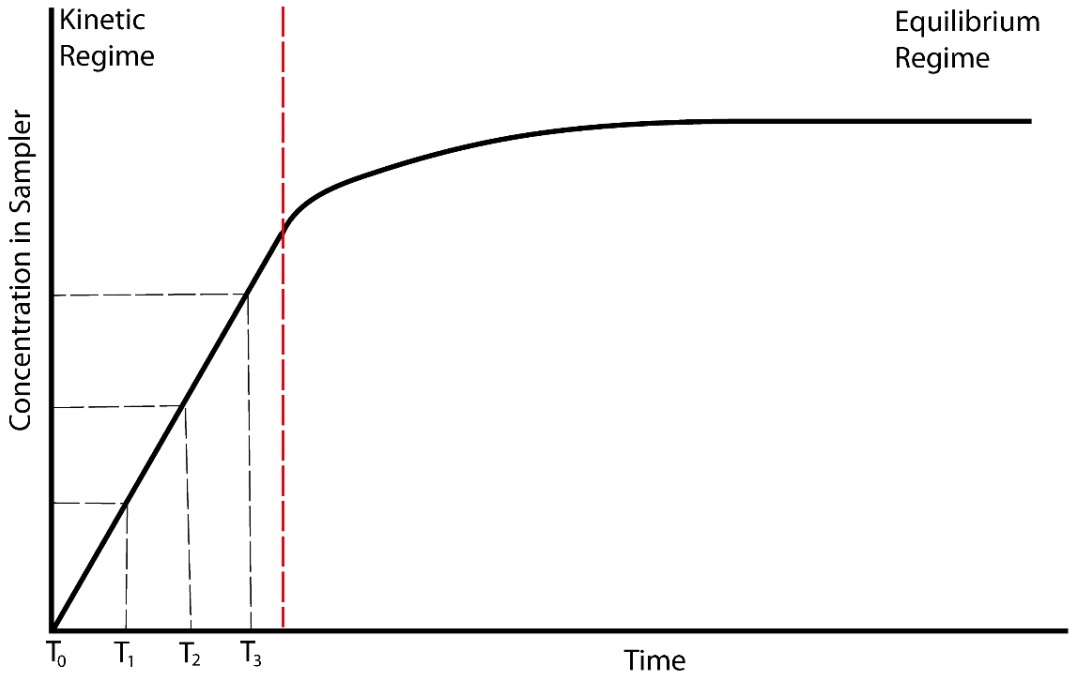


Figure 7.4. The theoretical accumulation of compounds onto a passive sampler overtime

7.4.1.1 Pilot study of passive sampling at Chew Stoke sewage treatment works

Chew Stoke sewage treatment works was chosen to develop this untargeted passive sampling method as discussed in Chapter 3 this riverine environment is controlled using lake water providing a consistent upstream chemical composition. The passive samplers used contained about 1.0 g of Oasis HLB solid phase per sampler housed between two pre-combusted glass fibre filters. Four passive samplers were deployed upstream and downstream of the sewage outfall and were removed after 7, 14, 21 and 28 days from the two sites. Four POCIS disks were stored in HPLC grade water as a comparative blank and one was extracted alongside each

time point. The filter of the upstream passive sampler collected after 28 days failed while deployed in the river and therefore, no extract was obtained for this time point.

The solid phase recovered from the disks was washed with HPLC grade water to remove any inorganic salts and dried under nitrogen. The solid phase was eluted with HPLC grade methanol (3 x 10 ml) and the extract dried under nitrogen. The solid phase was then dried at 50 ºC until a constant weight was reached to determine the amount of solid phase recovered from each of the samplers.

All passive sampler extracts were dissolved in methanol/water (*v/v*, 2000 µl). Comparative grab samples were collected as described in Section 2.1 and extracted as described in Section 2.2. The dried grab sample extracts were dissolved in methanol (700 µl). An aliquot (50 µl) was removed from each extract for DOC analysis which was dried using nitrogen and dissolved in water (20 ml, Milli q). An additional filtered water sample from each time point was also reserved for subsequent DOC analysis. DOC analysis was carried out as described in Section 2.3.1. All extracts were stored at -85 ºC for HPLC-MS analysis which was carried out as described in Section 2.5.3.

7.4.1.2 DOC

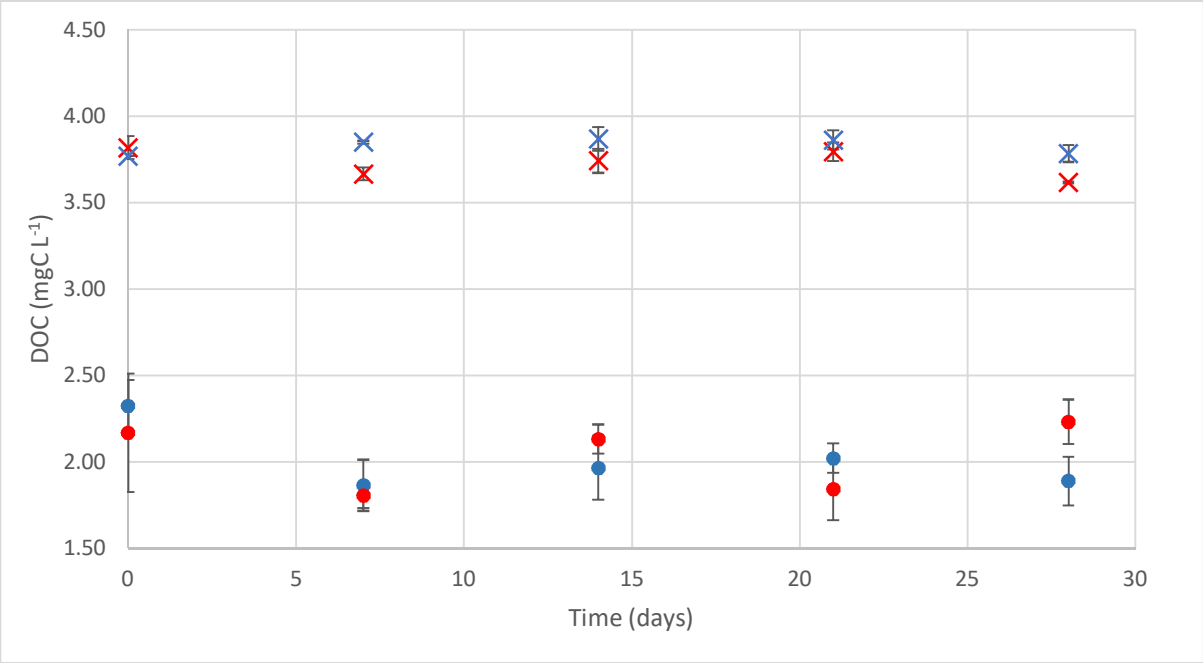


Figure 7.5. DOC of the filtered water samples collected upstream (x) and downstream (x) of the sewage outfall. The concentration of carbonaceous material in the grab sample SPE extracts from the upstream (●) and downstream (●) water samples.

Figure 7.5 shows the concentration of organic carbon found in the filtered water samples collected upstream and downstream of the sewage outfall at Chew Stoke and the concentration of carbon in the grab sample extracts. The amount of organic carbon in the grab sample SPE extracts was calculated using Equation 2.1. The graph clearly shows that the amount of DOM in the river and the amount of organic carbon in the grab sample extracts remains consistent over the 28 days that the passive samplers were deployed.

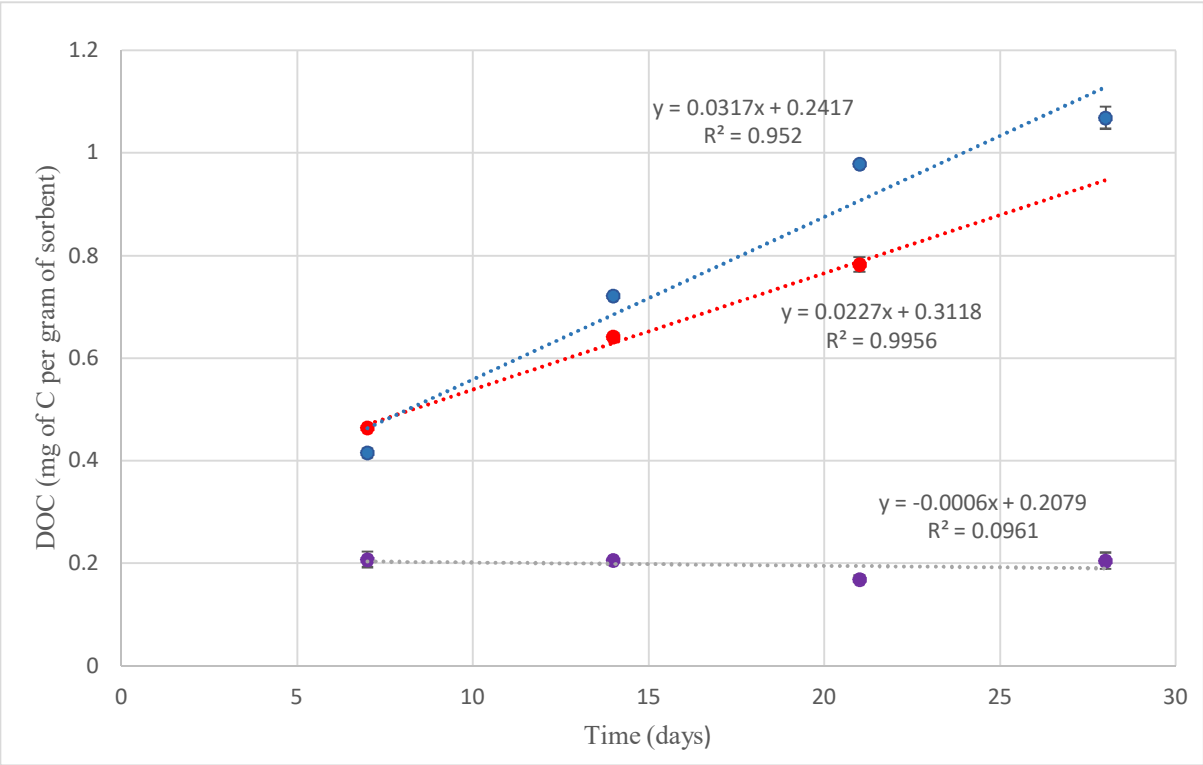


Figure 7.6. DOC in the passive samplers corrected for the mass of sorbent recovered from each of the samplers placed upstream (●) and downstream (●) of the sewage outfall, and the organic carbon recovered from the blank sampler (●).

The DOC extracted from the passive samplers was calculated using Equation 2.1. This value was corrected for the mass of solid phase recovered using Equation 7.1; where DOC_{ps} is the concentration of organic carbon per gram of solid phase recovered, DOC_{ex} is the concentration of organic carbon in the extracts and M_{sp} is the mass of the solid phase recovered from each sampler. Figure 7.6 shows the amount of organic carbon extracted per gram of sorbent recovered from the passive samplers. There is a clear linear increase in the amount of organic carbon in the riverine passive samplers over time. In contrast the amount of organic carbon in the blank remains constant. The concentration of organic carbon collected by the downstream

sampler has a steeper gradient when compared to the upstream. This indicated that there are additional compounds accumulating on the downstream sorbent despite the amount of organic carbon in the filtered water samples remaining constant over time. This suggests that there are additional organics input by the sewage outfall not found upstream which are binding to the downstream passive sampler.

$$DOC_{ps} = \frac{DOC_{ex}}{M_{sp}} \qquad \textbf{(7.2)}$$

7.4.1.3 HPLC-MS analysis of the passive sampler extracts

All of the extracts from the passive samplers were analysed using HPLC-MS as described in Section 2.3.3. As shown in Chapters 4, 5 and 6 the large amount of co-elution in the TICs made a meaningful comparison not possible. Peak picking and alignment were carried out as described 2.3.3.1. The peak area was corrected for the mass of the sorbent recovered.

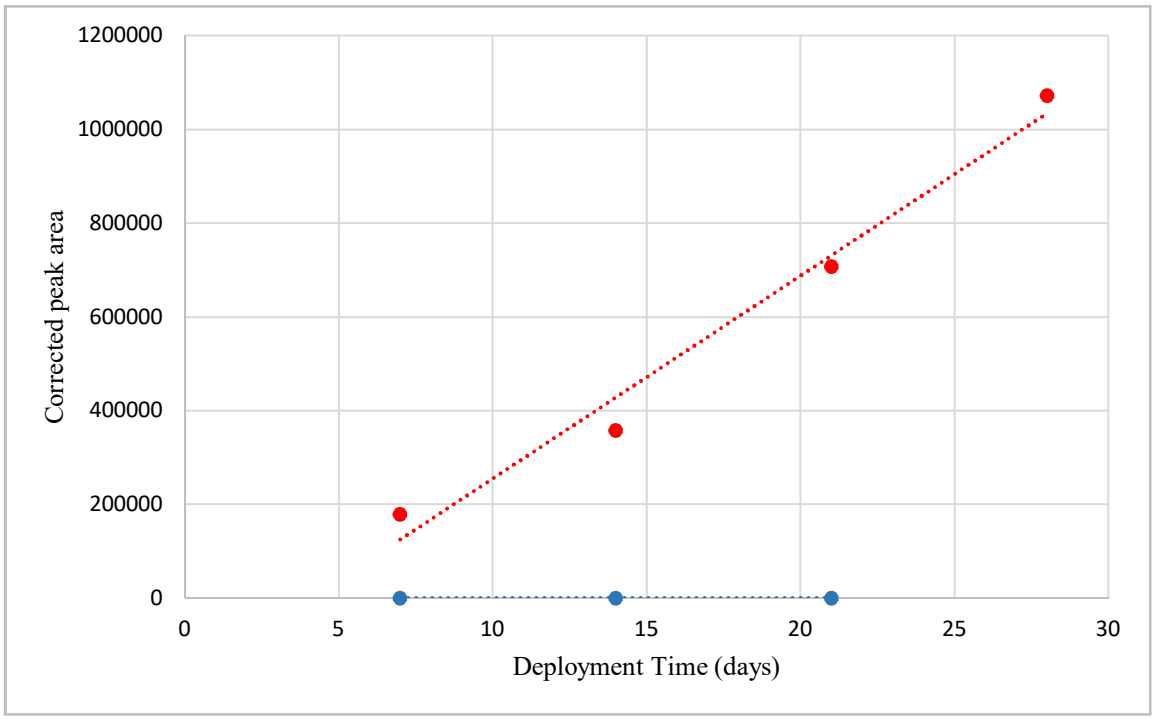


Figure 7.7. The corrected peak area for the ion *m/z* 325.1723 at 9.66 min plotted against the deployment time of the passive sampler for the upstream (●) and downstream (●) of the sewage treatment works.

To determine which compounds were coming from the sewage outfall, peaks detected in the passive samplers in the blank and upstream extracts were removed. Compounds with a concentration found to be linearly accumulating with time in the downstream sampler were retained for further analysis. This was done as compounds whose concentration did not increase

over time would not be viable for further passive sampling studies. Furthermore, these compounds would not explain the trend seen in the DOC analysis of the passive samplers. This was carried out by removing peaks which did not have a positive correlation when comparing the peak areas found in the EICs and had an $R^2 > 0.9$. In total 331 peaks were found in the downstream sampler which linearly accumulated with time and had an $R^2 > 0.9$. These 331 compounds were compiled into a target mass list and the product ion spectra collected as described in Section 2.3.4.

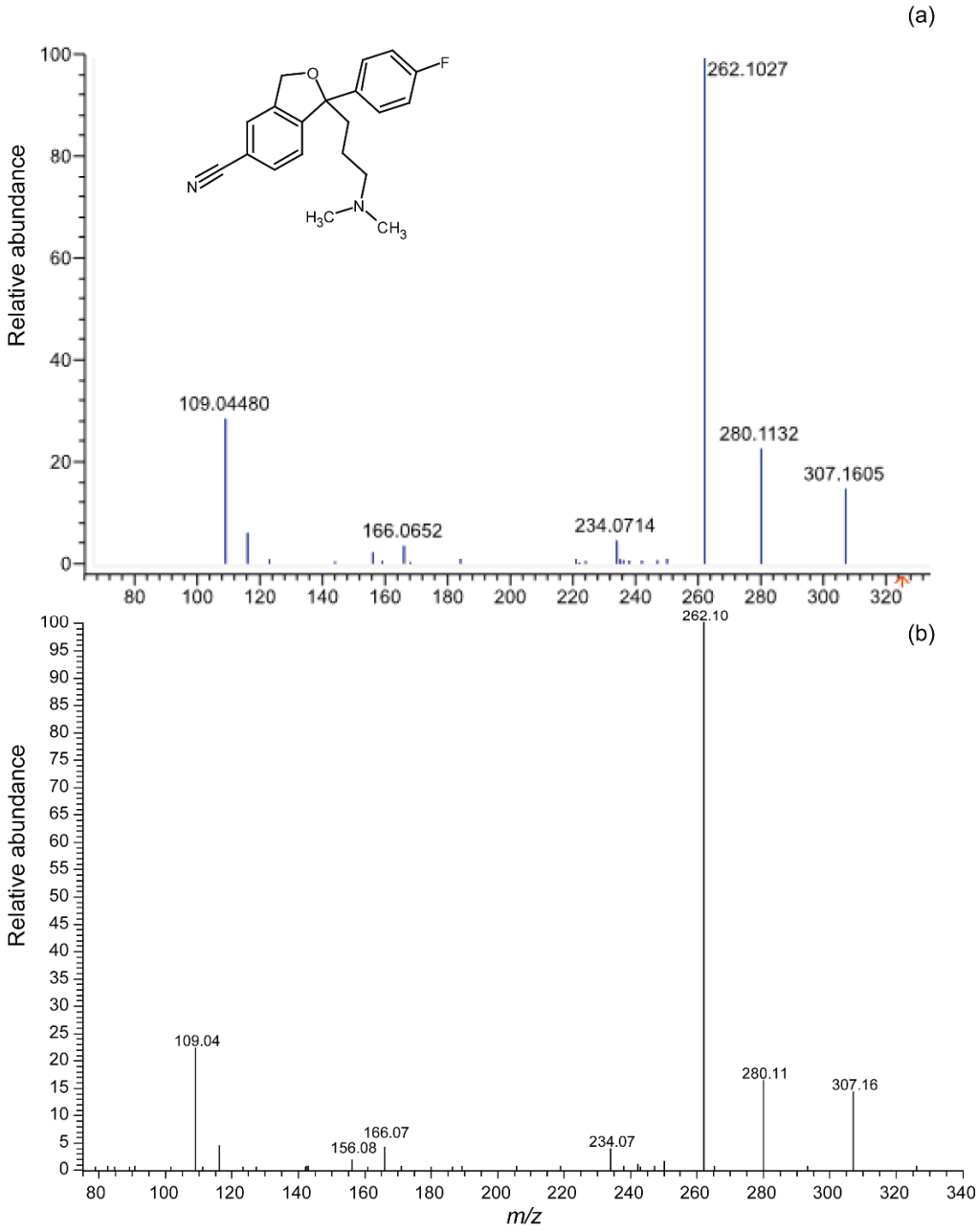


Figure 7.6. (a) mzCloud database recorded product ions spectrum for the precursor *m/z* 325.1711 CID 35 eV. (b) Recorded product ion spectrum of the ion *m/z* 325.1723 at 9.66 minutes CID 30 eV.

Figure 7.7 shows an example of one of the 331 compounds for which the concentration was found to increase linearly with time, the ion *m/z* 325.1723. The graph shows a clear linear increase in the concentration of the compound over time in the downstream passive sampler which has a R^2 of 0.98. In contrast, the upstream passive sampler shows no accumulation of this compound. This demonstrates that this compound is originating from the sewage outfall as it is only found in the downstream sampler.

Figure 7.6 shows the MS^2 of the ion *m/z* 325.1723 which was an exact match to the database product ion spectrum of citalopram. This compound was also identified in Chapter 4 when analysing Chew Stoke sewage treatment works. Citalopram as discussed in Chapter 4 is an antidepressant which has a known environmental impact (Henry *et al.*, 2004; Fong and Hoy, 2012).

In summary, this preliminary study shows that extracts from passive samplers can be used with untargeted analysis to determine which compounds are originating from a point source. This study has shown that the concentration of organic carbon on passive samplers deployed in the river linearly accumulates with time. This shows that both samplers were in the kinetic regime while deployed. The passive samplers clearly showed the presence of the sewage outfall using the DOC analysis as the downstream samplers accumulated more organic carbon at a faster rate than the upstream samplers. Analysing the passive sampler extracts using HPLC-MS allowed the additional compounds found in the downstream sampler to be differentiated when compared to the composition of the upstream POCIS extract. Compounds whose concentration increased linearly with time were targeted and identified. This shows the potential for these compounds to be used in future passive sampling studies.

To further this work, it would be beneficial to:

(i) Continue to identify the compounds (331) which linearly accumulated in the downstream passive sampler. As these are potentially pollutants from the sewage treatment works.

(ii) Acquire standards for the compounds identified and retrospectively calculate the rate of uptake of identified compounds onto the passive samplers using a laboratory-based study. This value is required to calculate the time weighted average as stated in equation 7.1.

(iii) Quantify the concentration of the compounds in the comparative grab samples and see if the calculated time weighted average is similar to the mean of the concentration in the grab samples collected

(iv) Compare the laboratory calculated rate (ii) to the accumulation rate observed in the passive samplers used in this study and calculate the theoretical time weighted average concentration in the river (iii). This will provide an interesting comparison to determine if the time weighted average concentration concurs with the grab sample concentrations.

This work demonstrates the potential value of using passive samplers with an untargeted workflow for a more comprehensive analysis of DOM.

7.4.2 Investigation of different extraction methods and analyses to further investigate the complexity of DOM

One of the limitations of this study and previous studies is that DOM needs to be extracted prior to analysis. No extraction method extracts all the DOM from a water sample, as discussed in the introduction. Therefore only a fraction of the overall DOM is studied (e.g. D'Archivio *et al.*, 2007; Green *et al.*, 2014; Minor *et al.*, 2014; Li *et al.*, 2017). Furthermore, this study focused upon a single ionisation method and polarity for the analysis of the DOM extracted from three treatment works. To more comprehensively analyse DOM:

(i) A range of SPE sorbents could be used to extract different fractions of DOM.

(ii) Additional extracts and the ones collected in this thesis could be re-analysed in negative ion mode on the Orbitrap to determine if different compounds can be identified.

Using the methods developed as part of this thesis these data sets could be compared to determine if they are different in composition, and if there are additional compounds which are significant in differentiating between source and riverine DOM.

7.4.3 Incorporation of further databases and in silico fragmentation software

A limited number of compounds were identified of the ions which were determined to be statistically significant. However, these statistically significant ions are still important. To improve upon the identification of compounds:

(i) A wider range of databases could be interrogated for other known compounds.

(ii) *In silico* fragmentation prediction software could be used to aid the identification of compounds from the recorded product ion spectra.

7.4.4 Better understanding of the sources of variation in sewage effluent and riverine DOM chemistry

As shown by the comparisons of the DI-HRMS spectra between different sites and sewage treatment works, all extracts from each location are different. Even the extracts from the same catchment are dissimilar. If assumed there are three main factors driving the differences seen between treated effluent; the secondary treatment method, geographical location and population demographic. Multiple sewage treatment work effluent could be sampled based on the contrasting and comparable aspects of these three factors. It would be hypothesising that:

(i) The effluent produced using the same secondary treatment methods would be found to be more similar than those produced using different treatment methods.
(ii) The effluent produced across similar population demographics will be more similar regardless of geographic location.

This approach would allow the investigation of the spatial variability of the micropollutants in DOM and provide a better understanding on the variability of sewage effluent. This will result in a better understanding of the compounds being discharged and why in targeted analysis the absence/presence and concentrations are so variable (Loos *et al.*, 2012). Using the methods developed in this thesis, it should be possible to determine which sewage effluent compositions are most similar and different.

7.4.5 Understanding of point sources in the environment

A diverse range of point sources are continually discharged into the environment including industrial outfalls, domestic homes and storm drains. However, there has been little research into the nature of the compounds coming from these potentially unique sources of organic matter. With 70 % of industrial waste being discharged into rivers untreated in developing countries, this presents an environmental and analytical challenge for researchers, to characterise the composition of these sources. Unlike sewage treatment works which have been studied using targeted analysis, industrial outfalls and other point sources have been less well characterised. This is likely due to industries producing waste products which are unique to the industry discharging them into the environment making it difficult to predict the composition of the industrial waste.

The methodology developed as part of this thesis does not require the pre-determination of the analytes prior to analysis of a point source. Therefore, it be ideal to determine:

(i) If the composition of the DOM discharged by an industry is significantly different to that found in the environment.
(ii) If the extracts are significantly different, then which ions discriminate between the riverine and industrial effluent DOM.
(iii) Which compounds are statistically significant.
(iv) Investigate the environmental significance of the compounds to determine if further treatment methods are required to process the industrial effluent.

7.4.6 Understanding the degradation of PPG

PPG was found at both Chew Stoke and Llanrwst and formed a large proportion of the compounds identified from the outfalls at these sewage treatment works. However, PPG was not found at Betws-y-Coed. PPG is known to biodegradable and as a result theoretically should not be identified at treatment works which use a secondary treatment process (Zgoła-Grześkowiak *et al.*, 2006; Hu *et al.*, 2008). Therefore, by investigating PPG using laboratory-based incubation experiments comparable to trickle filtration and active sludge treatment methods, the degradation of this oligomeric material could be further investigated. This could provide insight into:

(i) The identities of PPG metabolites
(ii) The rate of degradation of PPG using different treatment methods

7.4.7 Understanding the temporal variability of DOM

Previous studies using untargeted analysis have focused on the spatial variability of DOM between different water bodies and within a catchment (Dubinenkov *et al.*, 2015). However, non-specific monitoring techniques have shown that there are diurnal and seasonal changes in the concentration and composition of riverine DOM (Huber *et al.*, 1994; Lundquist and Cayan, 2002; Palmer-Felgate *et al.*, 2008; Carey and Migliaccio, 2009; Yates and Johnes, 2013). Recently, targeted analysis of the concentration of illicit drugs, antihistamines and antidepressants, have shown that their concentrations change in wastewater and surface water either diurnally, weekly, or seasonally (Karolak *et al.*, 2010; Nelson *et al.*, 2011; Golovko *et al.*, 2014). This is because the usage of a pharmaceutical agent is often related to an external cause. For example, anti-histamine use increases during spring and summer due to increases in

pollen levels, and their concentration in sewage correlates with use. Therefore, understanding the temporal variability of DOM in rivers and waste water would provide a better understanding of whether:

(i) Changes observed in non-specific monitoring techniques are due to a quantitative or qualitative change in the DOM.
(ii) The composition of sewage effluent or riverine DOM significantly change over time?
(iii) Compounds changing in abundance relate to climatic, seasonal, or diurnal factors, such as temperature, irradiance, seasonally determined human demand for particularly medicines or industrial products, e.g. agrochemical.

A future study could provide insight into whether the sampling frequency of water or effluent is representative of a source or river.

7.5 Concluding remarks

Overall, this thesis presents a novel statistically driven method for the untargeted HRMS analysis of point source DOM in the riverine environment. DI-HRMS of SPE extracts provide a qualitative screening method to determine if there is a significant difference between different DOM sources. If a significant difference is found, then Kruskal-Wallis analysis can be used to prioritise discriminating ions for identification. This provides a necessary data reduction step to determine which compounds to identify. The combination of DI-HRMS and HPLC-MS provides complimentary analysis methods which allow the identification of unknowns within the DOM extracts.

This novel method has the potential to be used for a wide range of applications, extending and complementing the current untargeted and targeted approaches to analysing DOM using MS.

References

Ahmad, S. R., & Reynolds, D. M. (1999). Monitoring of water quality using fluorescence technique: prospect of on-line process control. *Water Research*, **33**(9), 2069–2074.

Ahrens, L., Daneshvar, A., Lau, A. E., & Kreuger, J. (2015). Characterization of five passive sampling devices for monitoring of pesticides in water. *Journal of Chromatography A*, **1405**, 1–11.

Aitkenhead, J. A., Hope, D., & Billett, M. F. (1999). The relationship between dissolved organic carbon in stream water and soil organic carbon pools at different spatial scales. *Hydrological Processes*, **13**(8), 1289–1302.

Allard, B., Borén, H., Pettersson, C., & Zhang, G. (1994). Degradation of humic substances by UV irradiation. *Environment International*, **20**(1), 97–101.

Allard, S., Charrois, J. W. A., Joll, C. A., & Heitz, A. (2012). Simultaneous analysis of 10 trihalomethanes at nanogram per liter levels in water using solid-phase microextraction and gas chromatography mass-spectrometry. *Journal of Chromatography A*, **1238**, 15–21.

Almeida, Â., Calisto, V., Esteves, V. I., Schneider, R. J., Soares, A. M. V. M., Figueira, E., & Freitas, R. (2017). Ecotoxicity of the antihistaminic drug cetirizine to Ruditapes philippinarum clams. *Science of the Total* Environment, **601–602**, 793–801.

Andresen, J. A., Grundmann, A., & Bester, K. (2004). Organophosphorus flame retardants and plasticisers in surface waters. *Science of the Total Environment*, **332**(1–3), 155–166.

Arellano, A. R., Bianchi, T. S., Hutchings, J. A., Shields, M. R., & Cui, X. (2018). Differential effects of solid-phase extraction resins on the measurement of dissolved lignin-phenols and organic matter composition in natural waters. *Limnology and Oceanography: Methods*, **16**(1), 22–34.

Arias-Estévez, M., López-Periago, E., Martínez-Carballo, E., Simal-Gándara, J., Mejuto, J. C., & García-Río, L. (2008). The mobility and degradation of pesticides in soils and the pollution of groundwater resources. Agriculture, Ecosystems and Environment, **123**(4), 247–260.

AVERT. (2016). HIV and AIDS in the United Kingdom (UK). Retrieved July 14, 2018, from https://www.avert.org/professionals/hiv-around-world/western-central-europe-north-america/uk

Bajad, S., Khajuria, R. K., Suri, O. P., & Bedi, K. L. (2003). Characterisation of a new minor urinary metabolite of piperine, an omnipresent food component, by LC-MS/MS. *Journal of Separation Science,* **26**(9–10), 943–946.

Baker, A., & Spencer, R. G. M. (2004). Characterization of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy. *Science of the Total Environment*, **333**(1–3), 217–232.

Baker, D. R., Barron, L., & Kasprzyk-Hordern, B. (2014). Illicit and pharmaceutical drug consumption estimated via wastewater analysis. Part A: Chemical analysis and drug use estimates. *Science of the Total Environment*, **487**(1), 629–641.

Barkham, P. (2018, June 21). Cocaine in rivers harming endangered eels, study finds | Environment | The Guardian. The Guardian. Retrieved from https://www.theguardian.com/environment/2018/jun/21/cocaine-in-rivers-harming-endangered-eels-study-finds

Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., & Tranvik, L. J. (2009). The boundless carbon cycle. *Nature Geoscience*, **2**(9), 598–600.

Beggs, K. M. H., Summers, R. S., & McKnight, D. M. (2009). Characterizing chlorine oxidation of dissolved organic matter and disinfection by-product formation with fluorescence spectroscopy and parallel factor analysis. *Journal of Geophysical Research*, **114(**4), G04001

Beketov, M. A., Kefford, B. J., Schafer, R. B., & Liess, M. (2013). Pesticides reduce regional biodiversity of stream invertebrates. *Proceedings of the National Academy of Sciences*, **110**(27), 11039–11043.

Benton, H. P., Want, E. J., & Ebbels, T. M. D. (2010). Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *Bioinformatic*s, **26**(19), 2488–2489.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Wiswedel, B. (2009). KNIME - the Konstanz information miner. *ACM SIGKDD Explorations Newsletter*, **11**(1), 26.

Bourgeois, W., Burgess, J. E., & Stuetz, R. M. (2001). On-line monitoring of wastewater quality: A review. *Journal of Chemical Technology and Biotechnology*, **76**(4), 337–348.

Boyd-Boland, A. A., Magdic, S., & Pawliszyn, J. B. (1996). Simultaneous determination of 60 pesticides in water using solid-phase microextraction and gas chromatography–mass spectrometry. *The Analyst***, 121**(7), 929–937.

Brailsford, F. L., Glanville, H. C., Marshall, M. R., Golyshin, P. N., Johnes, P. J., Yates, C. A., … Jones, D. L. (2017). Microbial use of low molecular weight DOM in filtered and unfiltered freshwater: Role of ultra-small microorganisms and implications for water quality monitoring. *Science of the Total Environment*, **598**, 377–384.

Bringmann, G., Messer, K., Wohlfarth, M., Kraus, J., Dumbuya, K., & Rückert, M. (1999). HPLC-CD on-line coupling in combination with HPLC-NMR and HPLC-MS/MS for the determination of the full absolute stereostructure of new metabolites in plant extracts. *Analytical Chemistry*, **71**(14), 2678–2686.

Bristol Water. (2014). Spawn to be Wild | Bristol Water. Retrieved July 14, 2018, from https://www.bristolwater.co.uk/about-us/our-campaigns/spawn-to-be-wild/

Buszewski, B., & Szultka, M. (2012). Past, Present, and Future of Solid Phase Extraction: A Review. *Critical Reviews in Analytical Chemistry*, **42**(3), 198–213.

Cao, D., Huang, H., Hu, M., Cui, L., Geng, F., Rao, Z., … Kang, Y. (2015). Comprehensive characterization of natural organic matter by MALDI- and ESI-Fourier transform ion cyclotron resonance mass spectrometry. *Analytica Chimica Acta*, **866**, 48–58.

Capaldo, A., Gay, F., Lepretti, M., Paolella, G., Martucciello, S., Lionetti, L., Laforgia, V. (2018). Effects of environmental cocaine concentrations on the skeletal muscle of the European eel (Anguilla anguilla). *Science of the Total Environment*, **640–641**, 862–873.

Carey, R. O., & Migliaccio, K. W. (2009). Contribution of wastewater treatment plant effluents to nutrient dynamics in aquatic systems. *Environmental Management*.**44**(2), 205-217.

Carstea, E. M., Baker, A., Bieroza, M., Reynolds, D. M., & Bridgeman, J. (2014). Characterisation of dissolved organic matter fluorescence properties by PARAFAC analysis and thermal quenching. *Water Research*, **61**(0), 152–161.

Carstea, E. M., Bridgeman, J., Baker, A., & Reynolds, D. M. (2016). Fluorescence spectroscopy for wastewater monitoring: A review. *Water Research*, **95**, 205–219.

Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., … Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, **30**(10), 918–920.

Chen, H., Stubbins, A., Perdue, E. M., Green, N. W., Helms, J. R., Mopper, K., & Hatcher, P. G. (2014). Ultrahigh resolution mass spectrometric differentiation of dissolved organic matter isolated by coupled reverse osmosis-electrodialysis from various major oceanic water masses. *Marine Chemistry*, **164**, 48–59.

Chen, R., Yu, X., & Li, L. (2002). Characterization of poly(ethylene glycol) esters using low energy collision-induced dissociation in electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry*, **13**(7), 888–897.

Chen, Y., Wang, C., & Wang, Z. (2005). Residues and source identification of persistent organic pollutants in farmland soils irrigated by effluents from biological treatment plants. *Environment International*, **31**(6), 778–783.

Chopra, A. K., Sharma, M. K., & Chamoli, S. (2011). Bioaccumulation of organochlorine pesticides in aquatic system-an overview. *Environmental Monitoring and Assessment*, **173**(1–4), 905–916.

Corilo, Y. E., Vaz, B. G., Simas, R. C., Lopes Nascimento, H. D., Klitzke, C. F., Pereira, R. C. L., Eberlin, M. N. (2010). Petroleomics by EASI(±) FT-ICR MS. *Analytical Chemistry*, **82(**10), 3990–3996.

Cortés-Francisco, N., & Caixach, J. (2013). Molecular Characterization of Dissolved Organic Matter through a Desalination Process by High Resolution Mass Spectrometry. *Environmental Science & Technology*, **47**(17), 9619–9627.

Cortés-Francisco, N., Harir, M., Lucio, M., Ribera, G., Martínez-Lladó, X., Rovira, M., … Caixach, J. (2014). High-field FT-ICR mass spectrometry and NMR spectroscopy to characterize DOM removal through a nanofiltration pilot plant. *Water Research*, **67**, 154–165.

D'Andrilli, J., Cooper, W. T., Foreman, C. M., & Marshall, A. G. (2015). An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Communications in Mass Spectrometry*, **29**(24), 2385–2401.

D'Archivio, A. A., Fanelli, M., Mazzeo, P., & Ruggieri, F. (2007). Comparison of different sorbents for multiresidue solid-phase extraction of 16 pesticides from groundwater coupled with high-performance liquid chromatography. *Talanta*, **71**(1), 25–30.

Dalzell, B. J. J., Minor, E. C. C., & Mopper, K. M. M. (2009). Photodegradation of estuarine dissolved organic matter: a multi-method assessment of DOM transformation. *Organic Geochemistry*, **40**(2), 243–257.

De Haan, H., & De Boer, T. (1987). Applicability of light absorbance and fluorescence as measures of concentration and molecular size of dissolved organic carbon in humic Lake Tjeukemeer. *Water Research*, **21**(6), 731–734.

De Haan, H., De Boer, T., Kramer, H. A., & Voerman, J. (1982). Applicability of light absorbance as a measure of organic carbon in humic lake water. *Water Research*, **16**(6), 1047–1050.

DEFRA. (2002). UK Implementation of the EC Urban Waste Water Treatment Directive. Water Services, **20**. Retrieved from http://www.defra.gov.uk/publications/files/pb6655-uk-sewage-treatment-020424.pdf

DEFRA. (2012). Waste water treatment in the United Kingdom. Defra, **49**. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69592/pb13811-waste-water-2012.pdf

Degens, E. T., Kempe, S., & Richey, J. E. (1991). Biogeochemistry of major world rivers. In SCOPE 42: biogeochemistry of major world rivers. **42**, 353.

Dobbs, R. A., Wise, R. H., & Dean, R. B. (1972). The use of ultra-violet absorbance for monitoring the total organic carbon content of water and wastewater. *Water Research*, **6**(10), 1173–1180.

Drinking Water Inspectorate. (2001). Drinking Water Annual Report 2001. DEFRA.

Drinking Water Inspectorate. (2018). Guidance on implementing the Water Supply (Water Quality) Regulations in England and Wales. Drinking Water Inspectorate. Retrieved from http://www.dwi.gov.uk/stakeholders/guidance-and-codes-of-practice/wswq/index.html

Dubinenkov, I., Flerus, R., Schmitt-Kopplin, P., Kattner, G., & Koch, B. P. (2015). Origin-specific molecular signatures of dissolved organic matter in the Lena Delta. *Biogeochemistry*, **123**(1–2), 1–14.

Dvořák, L., Lederer, T., Jirků, V., Masák, J., & Novák, L. (2014). Removal of aniline, cyanides and diphenylguanidine from industrial wastewater using a full-scale moving bed biofilm reactor. *Process Biochemistry*, **49**(1), 102–109.

EEC, & European Union. Urban Waste Water directive. Council Directive (91/271/EEC), 34 Official J. Eur. Commun. L 1–16 (1991).

Emmett, B. A., Cooper, D., Smart, S., Jackson, B., Thomas, A., Cosby, B., Jones, D. L. (2016). Spatial patterns and environmental constraints on ecosystem services at a catchment scale. *Science of the Total Environment*, **572**(7), 1586–1600.

Environment Canada-Health Canada. (2013). Screening Assessment for the Challenge Diphenylguanidine.

Farkas, K., Cooper, D. M., McDonald, J. E., Malham, S. K., de Rougemont, A., & Jones, D. L. (2018). Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters. *Science of The Total Environment*, **634**, 1174–1183.

Fatta-Kassinos, D., Meric, S., & Nikolaou, A. (2011). Pharmaceutical residues in environmental waters and wastewater: Current state of knowledge and future research. *Analytical and Bioanalytical Chemistry*. **399**(1), 251-275.

Fellman, J. B., Hood, E., & Spencer, R. G. M. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review, *Limnology and Oceanography,* **55**(6) 2452–2462.

Flerus, R., Koch, B. P., Schmitt-Kopplin, P., Witt, M., & Kattner, G. (2011). Molecular level investigation of reactions between dissolved organic matter and extraction solvents using FT-ICR MS. *Marine Chemistry*, **124**(1–4), 100–107.

Fong, P. P., & Hoy, C. M. (2012). Antidepressants (venlafaxine and citalopram) cause foot detachment from the substrate in freshwater snails at environmentally relevant concentrations. *Marine and Freshwater Behaviour and Physiology*, **45**(2), 145–153.

Ford, A. T., & Fong, P. P. (2016). The effects of antidepressants appear to be rapid and at environmentally relevant concentrations. *Environmental Toxicology and Chemistry*, **35**(4), 794–798.

Franke, S., Meyer, C., Heinzel, N., Gatermann, R., Hühnerfuss, H., Rimkus, G., … Francke, W. (1999). Enantiomeric composition of the polycyclic musks HHCB and AHTN in different aquatic species. *Chirality*, **11**(10), 795–801.

Fritzmann, C., Löwenberg, J., Wintgens, T., & Melin, T. (2007). State-of-the-art of reverse osmosis desalination. *Desalination*, **216**(1–3), 1–76.

Gallampois, C. M. J., Schymanski, E. L., Bataineh, M., Buchinger, S., Krauss, M., Reifferscheid, G., & Brack, W. (2013). Integrated biological-chemical approach for the isolation and selection of polyaromatic mutagens in surface waters. *Analytical and Bioanalytical Chemistry*, **405**(28), 9101–9112.

Gavrilescu, M., Demnerová, K., Aamand, J., Agathos, S., & Fava, F. (2015). Emerging pollutants in the environment: Present and future challenges in biomonitoring, ecological risks and bioremediation. *New Biotechnology*, **32**(1), 147–156.

Goel, G., & Kaur, S. (2012). A Study on Chemical Contamination of Water Due to Household Laundry Detergents. *Journal of Human Ecology*, **38**(1), 65–69.

Golovko, O., Kumar, V., Fedorova, G., Randak, T., & Grabic, R. (2014). Seasonal changes in antibiotics, antidepressants/psychiatric drugs, antihistamines and lipid regulators in a wastewater treatment plant. *Chemosphere*, **111**, 418–426.

Gonsior, M., Schmitt-Kopplin, P., Stavklint, H., Richardson, S. D., Hertkorn, N., & Bastviken, D. (2014). Changes in dissolved organic matter during the treatment processes of a drinking water plant in sweden and formation of previously unknown disinfection byproducts. *Environmental Science and Technology,* **48**(21), 12714–12722.

Gonsior, M., Zwartjes, M., Cooper, W. J., Song, W., Ishida, K. P., Tseng, L. Y., Schmitt-Kopplin, P. (2011). Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. *Water Research*, **45**(9), 2943–2953.

Goslan, E. H., Krasner, S. W., Bower, M., Rocks, S. A., Holmes, P., Levy, L. S., & Parsons, S. A. (2009). A comparison of disinfection by-products found in chlorinated and chloraminated drinking waters in Scotland. *Water Research*, **43**(18), 4698–4706.

Green, N. W., Perdue, E. M., Aiken, G. R., Butler, K. D., Chen, H., Dittmar, T., Stubbins, A. (2014). An intercomparison of three methods for the large-scale isolation of oceanic dissolved organic matter. *Marine Chemistry*, **161**, 14–19.

Gros, M., Petrović, M., & Barceló, D. (2006). Development of a multi-residue analytical methodology based on liquid chromatography-tandem mass spectrometry (LC-MS/MS) for screening and trace level determination of pharmaceuticals in surface and wastewaters. *Talanta*, **70**(4), 678–690.

Gros, M., Petrović, M., & Barceló, D. (2009). Tracing pharmaceutical residues of different therapeutic classes in environmental waters by using liquid chromatography/quadrupole-linear ion trap mass spectrometry and automated library searching. *Analytical Chemistry*, **81**(3), 898–912.

Haddad, T., & Kümmerer, K. (2014). Characterization of photo-transformation products of the antibiotic drug Ciprofloxacin with liquid chromatography-tandem mass spectrometry in combination with accurate mass determination using an LTQ-Orbitrap. *Chemosphere*, **115**(1), 40–46.

Harman, C., Allan, I. J., & Vermeirssen, E. L. M. (2012). Calibration and use of the polar organic chemical integrative sampler-a critical review. *Environmental Toxicology and Chemistry*, **31**(12), 2724–2738.

Harman, C., Reid, M., & Thomas, K. V. (2011). In situ calibration of a passive sampling device for selected illicit drugs and their metabolites in wastewater, and subsequent year-long assessment of community drug usage. *Environmental Science and Technology*, **45**(13), 5676–5682.

Harris, B. D., Brown, T. A., McGehee, J. L., Houserova, D., Jackson, B. A., Buchel, B. C., Stenson, A. C. (2015). Characterization of Disinfection By-Products from Chromatographically Isolated NOM through High-Resolution Mass Spectrometry. *Environmental Science and Technology*, **49**(24), 14239–14248.

Heathwaite, A. L., Johnes, P. J., & Peters, N. E. (1996). Trends in nutrients. *Hydrological Processes*, **10**(2), 263–293.

Heberer, T. (2002). Occurrence, Fate, and Assessment of Polycyclic Musk Residues in the Aquatic Environment of Urban Areas — A Review. *Acta Hydrochimica et Hydrobiologica*, **30**(56), 227–243.

Henry, T. B., Kwon, J.-W., Armbrust, K. L., & Black, M. C. (2004). Acute and chronic toxicity of five selective serotonin reuptake inhibitors in Ceriodaphnia dubia. *Environmental Toxicology and Chemistry*, **23**(9), 2229–2233.

Hertkorn, N., Harir, M., Cawley, K. M., Schmitt-Kopplin, P., & Jaffé, R. (2016). Molecular characterization of dissolved organic matter from subtropical wetlands: A comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences*, **13**(8), 2257–2277.

Hertkorn, N., Harir, M., Koch, B. P., Michalke, B., & Schmitt-Kopplin, P. (2013). High-field NMR spectroscopy and FTICR mass spectrometry: Powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences*, **10**(3), 1583–1624.

Herzsprung, P., Hertkorn, N., von Tümpling, W., Harir, M., Friese, K., & Schmitt-Kopplin, P. (2014). Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Analytical and Bioanalytical Chemistry*, **406**(30), 7977–7987.

Herzsprung, P., von Tümpling, W., Wendt-Potthoff, K., Hertkorn, N., Harir, M., Schmitt-Kopplin, P., & Friese, K. (2017). High field FT-ICR mass spectrometry data sets enlighten qualitative DOM alteration in lake sediment porewater profiles. *Organic Geochemistry*, **108**, 51–60.

Hockaday, W. C., Purcell, J. M., Marshall, A. G., Baldock, J. A., & Hatcher, P. G. (2009). Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: A qualitative assessment. *Limnology and Oceanography: Methods*, **7**(1), 81–95.

Hollender, J., Schymanski, E. L., Singer, H., & Ferguson, P. L. (2017). Non-target screening with high resolution mass spectrometry in the environment: Ready to go? *Environmental Science & Technology*, **51**(20), 11505-11512.

Hu, X., Liu, X., Tani, A., Kimbara, K., & Kawai, F. (2008). Proposed Oxidative Metabolic Pathway for Polypropylene Glycol in Sphingobium sp. Strain PW-1. *Bioscience, Biotechnology, and Biochemistry*, **72**(4), 1115–1118.

Huang, Y. J., Qi, G. R., & Wang, Y. H. (2002). Controlled ring-opening polymerization of propylene oxide catalyzed by double metal-cyanide complex. *Journal of Polymer Science, Part A: Polymer Chemistry*, **40**(8), 1142–1150.

Huber, S. A., Balz, A., & Frimmel, F. H. (1994). Identification of diffuse and point sources of dissolved organic carbon (DOC) in a small stream (Alb, Southwest Germany), using gel filtration chromatography with high-sensitivity DOC-detection. *Fresenius' Journal of Analytical Chemistry*, **350**(7–9), 496–503.

Hug, C., Ulrich, N., Schulze, T., Brack, W., & Krauss, M. (2014). Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environmental Pollution*, **184**, 25–32.

Ishii, S. K. L., & Boyer, T. H. (2012). Behavior of reoccurring parafac components in fluorescent dissolved organic matter in natural and engineered systems: A critical review. *Environmental Science and Technology*, **46**(4), 2006–2017.

Jeong, Y., Schïäffer, A., & Smith, K. (2017). Equilibrium partitioning of organic compounds to OASIS HLB as a function of compound concentration, pH, temperature and salinity. *Chemosphere*, *174*, 297–305.

Jobling, S., Nolan, M., Tyler, C. R., Brighty, G., & Sumpter, J. P. (1998). Widespread sexual disruption in wild fish. *Environmental Science and Technology*, **32**(17), 2498–2506.

Jones, L., Ronan, J., McHugh, B., McGovern, E., & Regan, F. (2015). Emerging priority substances in the aquatic environment: A role for passive sampling in supporting WFD monitoring and compliance. *Analytical Methods*, **7**(19), 7976–7984.

Kaczala, F., & Blum, S. E. (2016). The Occurrence of Veterinary Pharmaceuticals in the Environment: A Review. *Current Analytical Chemistry*, **12**(3), 169–182.

Kaiser, E., Simpson, A. J., Dria, K. J., Sulzberger, B., & Hatcher, P. G. (2003). Solid-state and multidimensional solution-state NMR of solid phase extracted and ultrafiltered riverine dissolved organic matter. *Environmental Science and Technology*, **37**(13), 2929–2935.

Kamel, K. S., & Halperin, M. L. (2014). Metabolic Acidosis. *Urology at a Glance* (pp. 141–170).

Karolak, S., Nefau, T., Bailly, E., Solgadi, A., & Levi, Y. (2010). Estimation of illicit drugs consumption by wastewater analysis in Paris area (France). *Forensic Science International*, **200**(1–3), 153–160.

Kasprzyk-Hordern, B., Dinsdale, R. M., & Guwy, A. J. (2007). Multi-residue method for the determination of basic/neutral pharmaceuticals and illicit drugs in surface water by solid-phase

extraction and ultra performance liquid chromatography-positive electrospray ionisation tandem mass spectrometry. J*ournal of Chromatography A*, **1161**(1–2), 132–145.

Kawai, F., & Schink, B. (1987). The biochemistry of degradation of polyethers. *Critical Reviews in Biotechnology*, **6**(3), 273–307.

Kellerman, A. M., Dittmar, T., Kothawala, D. N., & Tranvik, L. J. (2014). Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nature Communications*, **5**(5), 1–8.

Kellerman, A. M., Kothawala, D. N., Dittmar, T., & Tranvik, L. J. (2015). Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nature Geoscience,* **8**(6), 454–457.

Kim, H. C., & Yu, M. J. (2007). Characterization of aquatic humic substances to DBPs formation in advanced treatment processes for conventionally treated water. *Journal of Hazardous Materials*, **143**(1–2), 486–493.

Kim, S., Simpson, A. J., Kujawinski, E. B., Freitas, M. A., & Hatcher, P. G. (2003). High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C18 solid phase disk. *Organic Geochemistry*, **34**(9), 1325–1335.

Kirwan, J. A., Broadhurst, D. I., Davidson, R. L., & Viant, M. R. (2013). Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Analytical and Bioanalytical Chemistry*, **405**(15), 5147–5157.

Kirwan, J. A., Weber, R. J. M., Broadhurst, D. I., & Viant, M. R. (2014). Direct infusion mass spectrometry metabolomics dataset: A benchmark for data processing and quality control. *Scientific Data*, 1, 140012.

Koch, B. P., Dittmar, T., Witt, M., & Kattner, G. (2007). Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Analytical Chemistry*, **79**(4), 1758–1763.

Kosonen, J., & Kronberg, L. (2009). The occurrence of antihistamines in sewage waters and in recipient rivers. *Environmental Science and Pollution Research*, **16**(5), 555–564.

Krasner, S. W., Weinberg, H. S., Richardson, S. D., Pastor, S. J., Chinn, R., Sclimenti, M. J., … Thruston, A. D. (2006). Occurrence of a new generation of disinfection by products. *Environmental Science and Technology*, **40**(23), 7175–7185.

Kujawinski, E. B., & Behn, M. D. (2006). Automated analysis of electrospray ionization fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Analytical Chemistry*, **78**(13), 4363–4373.

Lababidi, S., & Schrader, W. (2014). Online normal-phase high-performance liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry: Effects of different ionization methods on the characterization of highly complex crude oil mixtures. *Rapid Communications in Mass Spectrometry*, **28**(12), 1345–1352.

Lankadurai, B. P., Nagato, E. G., & Simpson, M. J. (2013). Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environmental Reviews*, **21**(3), 180–205.

Lefebvre, O., & Moletta, R. (2006, December 1). Treatment of organic pollution in industrial saline wastewater: A literature review. *Water Research*. 40(20), 3671-3682.

Li, H., & Minor, E. C. (2015). Dissolved organic matter in Lake Superior: insights into the effects of extraction methods on chemical composition. *Environ. Sci.: Processes Impacts,* **17**(10), 1829–1840.

Li, M. (2013). Acute toxicity of 30 pharmaceutically active compounds to freshwater planarians, Dugesia japonica. *Toxicological and Environmental Chemistry*, **95**(7), 1157–1170.

Li, Y., Harir, M., Lucio, M., Gonsior, M., Koch, B. P., Schmitt-Kopplin, P., & Hertkorn, N. (2016). Comprehensive structure-selective characterization of dissolved organic matter by reducing molecular complexity and increasing analytical dimensions. *Water Research*, **106**, 477–487.

Li, Y., Harir, M., Uhl, J., Kanawati, B., Lucio, M., Smirnov, K. S., … Hertkorn, N. (2017). How representative are dissolved organic matter (DOM) extracts? A comprehensive study of sorbent selectivity for DOM isolation. *Water Research*, **116**, 316–323.

Lindholm-Lehto, P. C., Ahkola, H. S. J., Knuutinen, J. S., Koistinen, J., Lahti, K., Vahtera, H., & Herve, S. H. (2016). Suitability of passive sampling for the monitoring of pharmaceuticals in Finnish surface waters. *Environmental Science and Pollution Research*, **23**(18), 18043–18054.

Liška, I. (2000). Fifty years of solid-phase extraction in water analysis - Historical development and overview. *Journal of Chromatography A*, **885**(1–2), 3–16.

Liu, J., Jiang, M., Li, G., Xu, L., & Xie, M. (2010). Miniaturized salting-out liquid-liquid extraction of sulfonamides from different matrices. *Analytica Chimica Acta*, **679**(1–2), 74–80.

Lloyd, C. E. M., Michaelides, K., Chadwick, D. R., Dungait, J. A. J., & Evershed, R. P. (2012). Tracing the flow-driven vertical transport of livestock-derived organic matter through soil using biomarkers. Organic Geochemistry, **43**(8), 56–66.

Longnecker, K., & Kujawinski, E. B. (2011). Composition of dissolved organic matter in groundwater. *Geochimica et Cosmochimica Acta*, **75**(10), 2752–2761.

Longwell, J., & Maniece, W. D. (1955). Determination of anionic detergents in sewage, sewage effluents and river waters. *The Analyst*, **80**(948), 167–171.

Loos, R., Carvalho, R., Comero, S., António, D., Ghiani, M., Lettieri, T., Gans, O. (2012). EU Wide Monitoring Survey on Waste Water Treatment Plant Effluents. *JRC scientific and policy report*. (p 1-138)

Loraine, G. A., & Pettigrove, M. E. (2006). Seasonal variations in concentrations of pharmaceuticals and personal care products in drinking water and reclaimed wastewater in Southern California. *Environmental Science and Technology*, **40**(3), 687–695.

Lundquist, J. D., & Cayan, D. R. (2002). Seasonal and Spatial Patterns in Diurnal Cycles in Streamflow in the Western United States. *Journal of Hydrometeorology*, **3**(5), 591–603.

Lv, J., Zhang, S., Luo, L., & Cao, D. (2016). Solid-phase extraction-stepwise elution (SPE-SE) procedure for isolation of dissolved organic matter prior to ESI-FT-ICR-MS analysis. *Analytica Chimica Acta*, **948**, 55–61.

Maida, C. M., Di Gaudio, F., Tramuto, F., Mazzucco, W., Piscionieri, D., Cosenza, A., & Viviani, G. (2017). Illicit drugs consumption evaluation by wastewater-based epidemiology in the urban area of Palermo city (Italy). *Annali Dell'Istituto Superiore Di Sanita*, **53**(3), 192–198.

Malaj, E., von der Ohe, P. C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P., Schäfer, R. B. (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proceedings of the National Academy of Sciences*, **111**(26), 9549–9554.

Mangal, V., Stock, N. L., & Guéguen, C. (2016). Molecular characterization of phytoplankton dissolved organic matter (DOM) and sulfur components using high resolution Orbitrap mass spectrometry. *Analytical and Bioanalytical Chemistry*, **408**(7), 1891–1900.

Marshall, A. G., & Rodgers, R. P. (2004). Petroleomics: The Next Grand Challenge for Chemical Analysis. *Accounts of Chemical Research*, **37**(1), 53–59.

Melendez-Perez, J. J., Martínez-Mejía, M. J., Awan, A. T., Fadini, P. S., Mozeto, A. A., & Eberlin, M. N. (2016). Characterization and comparison of riverine, lacustrine, marine and estuarine dissolved organic matter by ultra-high resolution and accuracy Fourier transform mass spectrometry. *Organic Geochemistry*, **101**, 99–107.

Melendez-Perez, J. J., Martínez-Mejia, M. J., & Eberlin, M. N. (2016). A reformulated aromaticity index equation under consideration for non-aromatic and non-condensed aromatic cyclic carbonyl compounds. *Organic Geochemistry*, **95**, 29–33.

Metcalfe, C., Hoque, M. E., Sultana, T., Murray, C., Helm, P., & Kleywegt, S. (2014). Monitoring for contaminants of emerging concern in drinking water using POCIS passive samplers. Environmental Science: Processes & Impacts, **16**(3), 473

Miao, X. S., Yang, J. J., & Metcalfe, C. D. (2005). Carbamazepine and its metabolites in wastewater and in biosolids in a municipal wastewater treatment plant. *Environmental Science and Technology*, **39**(19), 7469–7475.

Mills, G. A., Vrana, B., Allan, I., Alvarez, D. A., Huckins, J. N., & Greenwood, R. (2007). Trends in monitoring pharmaceuticals and personal-care products in the aquatic environment by use of passive sampling devices. Analytical and Bioanalytical Chemistry, **387**(4), 1153–1157.

Minor, E. C., Swenson, M. M., Mattson, B. M., & Oyler, A. R. (2014). Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci.: Processes Impacts,* **16**(9), 2064–2079.

Morin, N., Camilleri, J., Cren-Olivé, C., Coquery, M., & Miège, C. (2013). Determination of uptake kinetics and sampling rates for 56 organic micropollutants using “pharmaceutical” POCIS. *Talanta*, **109**, 61–73.

Mosher, J. J., Kaplan, L. A., Podgorski, D. C., McKenna, A. M., & Marshall, A. G. (2015). Longitudinal shifts in dissolved organic matter chemogeography and chemodiversity within headwater streams: a river continuum reprise. *Biogeochemistry*, **124(**1–3), 371–385.

Müller, M. B., Fritz, W., Lankes, U., & Frimmel, F. H. (2004). Ultrafiltration of Nonionic Surfactants and Dissolved Organic Matter. *Environmental Science and Technology*, **38**(4), 1124–1132.

Murray, K. E., Thomas, S. M., & Bodour, A. A. (2010). Prioritizing research for trace pollutants and emerging contaminants in the freshwater environment. *Environmental Pollution*. **158**(12), 3462-3471.

Nelson, E. D., Do, H., Lewis, R. S., & Carr, S. A. (2011). Diurnal variability of pharmaceutical, personal care product, estrogen and alkylphenol concentrations in effluent from a tertiary wastewater treatment facility. *Environmental Science and Technology*, **45**(4), 1228–1234.

Nikinmaa, M. (2014). An Introduction to Aquatic Toxicology.

Norrman, B. (1993). Filtration of water samples for DOC studies. *Marine Chemistry*, **41**(1–3), 239–242.

Novak, T. J., Grinberg, N., Hartman, B., Marcinko, S., DiMichele, L., & Mao, B. (2010). LCMS using a hybrid quadrupole time of flight mass spectrometer for impurity identification during process chemical development of a novel integrase inhibitor. *Journal of Pharmaceutical and Biomedical Analysis*, **51**(1), 78–83.

Novotny, V. (1999). Diffuse pollution from agriculture - A worldwide outlook. *Water Science and Technology*, **39**(3), 1–13.

Obach, R. S. (2013). Pharmacologically Active Drug Metabolites: Impact on Drug Discovery and Pharmacotherapy. *Pharmacological Reviews*, **65**(2), 578–640.

Obernosterer, I., & Benner, R. (2004). Competition between biological and photochemical processes in the mineralization of dissolved organic carbon. *Limnology and Oceanography,* **49**(1), 117–124.

Oetken, M., Nentwig, G., Löffler, D., Ternes, T., & Oehlmann, J. (2005). Effects of pharmaceuticals on aquatic invertebrates. Part I. The antiepileptic drug carbamazepine. *Archives of Environmental Contamination and Toxicology*, **49**(3), 353–361.

Okuno, S., Ohmoto, M., Arakawa, R., Okuno, S., Ohmoto, M., & Arakawa, R. (2003). Analysis of Polypropyleneglycols Using ESI-MS Analysis of polypropyleneglycols using electrospray ionization mass spectrometry. Effects of cationizing agents on the mass spectra. *Eur. J. Mass Spectrom*, **9**, 97–103.

Osborne, D. M., Podgorski, D. C., Bronk, D. A., Roberts, Q., Sipler, R. E., Austin, D., … Cooper, W. T. (2013). Molecular-level characterization of reactive and refractory dissolved natural organic

nitrogen compounds by atmospheric pressure photoionization coupled to Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry*, **27**(8), 851–858.

Palmer-Felgate, E. J., Jarvie, H. P., Williams, R. J., Mortimer, R. J. G., Loewenthal, M., & Neal, C. (2008). Phosphorus dynamics and productivity in a sewage-impacted lowland chalk stream. *Journal of Hydrology*, **351**(1–2), 87–97

Peček, G., Pavlović, D. M., & Babić, S. (2013). Development and validation of a SPE-GC-MS method for the determination of pesticides in surface water. *International Journal of Environmental Analytical Chemistry*, **93**(12), 1311–1328.

Petrie, B., Barden, R., & Kasprzyk-Hordern, B. (2014). A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. *Water Research*, **72**, 3–27.

Pfeiffer, E., Pavelescu, G., Baker, A., Roman, C., Ioja, C., & Savastru, D. (2008). Pollution analysis on the Arges River using fluorescence spectroscopy. *Journal of Optoelectronics and Advanced Materials*, **10**(6), 1489–1494.

Phungsai, P., Kurisu, F., Kasuga, I., & Furumai, H. (2016). Molecular characterization of low molecular weight dissolved organic matter in water reclamation processes using Orbitrap mass spectrometry. *Water Research*, **100**, 526–536.

Pinckney, J. L., Paerl, H. W., Tester, P., & Richardson, T. L. (2001). The role of nutrient loading and eutrophication in estuarine ecology. Environmental Health Perspectives, **109**, 699-709

Raeke, J., Lechtenfeld, O. J., Wagner, M., Herzsprung, P., & Reemtsma, T. (2016). Selectivity of solid phase extraction of freshwater dissolved organic matter and its effect on ultrahigh resolution mass spectra. *Environ. Sci.: Processes Impacts*, **18**(7), 918–927.

Reemtsma, T. (2009). Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry. Status and needs. *Journal of Chromatography A*, **1216**(18), 3687-3701.

Relyea, R. A. (2009). A cocktail of contaminants: How mixtures of pesticides at low concentrations affect aquatic communities. *Oecologia*, **159**(2), 363–376.

Reynolds, D. M., & Ahmad, S. R. (1997). Rapid and direct determination of wastewater BOD values using a fluorescence technique. *Water Research*, **31**(8), 2012–2018.

Rodil, R., Quintana, J. B., & Reemtsma, T. (2005). Liquid chromatography-tandem mass spectrometry determination of nonionic organophosphorus flame retardants and plasticizers in wastewater samples. *Analytical Chemistry*, **77**(10), 3083–3089.

Rodrigues, A. M., Ferreira, V., Cardoso, V. V., Ferreira, E., & Benoliel, M. J. (2007). Determination of several pesticides in water by solid-phase extraction, liquid chromatography and electrospray tandem mass spectrometry. *Journal of Chromatography A*, **1150**(1–2), 267–278.

Röst, H. L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., Kohlbacher, O. (2016). OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nature Methods*, **13**(9), 741-748.

Roth, V.-N., Dittmar, T., Gaupp, R., & Gleixner, G. (2015). The Molecular Composition of Dissolved Organic Matter in Forest Soils as a Function of pH and Temperature. *PLoS ONE*, **10**(3), e0119188.

Ruff, M., Mueller, M. S., Loos, M., & Singer, H. P. (2015). Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry - Identification of unknown sources and compounds. *Water Research*, **87**, 145–154.

Rujiralai, T. (2007). Detection and Quantification of Estrogens in the Environment. University of Bristol.

Rujiralai, T., Bull, I. D., Llewellyn, N., & Evershed, R. P. (2011). In situ polar organic chemical integrative sampling (POCIS) of steroidal estrogens in sewage treatment works discharge and river water. *Journal of Environmental Monitoring*, **13**(5), 1427–1434.

Rychłowska, J., Zgoła, A., Grześkowiak, T., & Łukaszewski, Z. (2003). Isolation of poly(propylene glycol)s from water for quantitative analysis by reversed-phase liquid chromatography. *Journal of Chromatography A*, **1021**(1–2), 11–17.

Sabik, H., Jeannot, R., & Rondeau, B. (2000). Multiresidue methods using solid-phase extraction techniques for monitoring priority pesticides, including triazines and degradation products, in ground and surface waters. *Journal of Chromatography A*, **885**(1–2), 217–236.

Sáenz Barrio, C., Romero Melgosa, E., Sanz Asensio, J., & Galbán Bernal, J. (1996). Extraction of pesticides from aqueous samples: A comparative study. *Mikrochimica Acta*, **122**(3–4), 267–277.

Santos, L. H. M. L. M., Gros, M., Rodriguez-Mozaz, S., Delerue-Matos, C., Pena, A., Barceló, D., & Montenegro, M. C. B. S. M. (2013). Contribution of hospital effluents to the load of pharmaceuticals in urban wastewaters: Identification of ecologically relevant pharmaceuticals. Science of the Total Environment, **461–462**, 302–316.

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environmental Science & Technology*, **48**(4), 2097–2098.

Schymanski, E. L., Singer, H. P., Longrée, P., Loos, M., Ruff, M., Stravs, M. A., Hollender, J. (2014). Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environmental Science & Technology*, **48**(3), 1811–1818.

Schymanski, E. L., Singer, H. P., Slobodnik, J., Ipolyi, I. M., Oswald, P., Krauss, M., Hollender, J. (2015). Non-target screening with high-resolution mass spectrometry: Critical review using a collaborative trial on water analysis. *Analytical and Bioanalytical Chemistry*, **407**(21), 6237–6255.

Segev, O., Kushmaro, A., & Brenner, A. (2009). Environmental impact of flame retardants (persistence and biodegradability). *International Journal of Environmental Research and Public Health*, **6**(2), 478–491.

Sella, A., & Lorna, S. (2014). How drugs are entering UK water systems through urine - BBC News. Retrieved September 25, 2018, from https://www.bbc.co.uk/news/health-29108330

Serkiz, S. M., & Perdue, E. M. (1990). Isolation of dissolved organic matter from the suwannee river using reverse osmosis. *Water Research*, **24**(7), 911–916.

Several, B., Water, E. U., Directive, F., Directive, D. W., State, E. U. M., Platform, I., … Norman, T. (2009). NORMAN Position Paper Collection , exchange and interpretation of data on emerging substances, 1–3. Retrieved from http://www.norman-network.net

Sheng, L.-H., Chen, H.-R., Huo, Y.-B., Wang, J., Zhang, Y., Yang, M., & Zhang, H.-X. (2014). Simultaneous Determination of 24 Antidepressant Drugs and Their Metabolites in Wastewater by Ultra-High Performance Liquid Chromatography–Tandem Mass Spectrometry. *Molecules*, **19**(1), 1212–1222.

Shiklomanov, I. (1993). World fresh water resources. Water in Crisis a Guide to the World's Fresh Water Resources.

Simjouw, J. P., Minor, E. C., & Mopper, K. (2005). Isolation and characterization of estuarine dissolved organic matter: Comparison of ultrafiltration and C18 solid-phase extraction techniques. *Marine Chemistry*, **96**(3–4), 219–235.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, **78**(3), 779–787.

Sonnenschein, C., & Soto, A. M. (1998). An updated review of environmental estrogen and androgen mimics and antagonists1. *The Journal of Steroid Biochemistry and Molecular Biology*, **65**(1–6), 143–150.

Spencer, R. G. M., Butler, K. D., & Aiken, G. R. (2012). Dissolved organic carbon and chromophoric dissolved organic matter properties of rivers in the USA. *Journal of Geophysical Research: Biogeosciences*, **117**(3).

Spitzy, A., & Leenheer, J. (1991). Dissolved Organic Carbon in Rivers. In Biogeochemistry of Major World Rivers (pp. 1–19).

Stackelberg, P. E., Furlong, E. T., Meyer, M. T., Zaugg, S. D., Henderson, A. K., & Reissman, D. B. (2004). Persistence of pharmaceutical compounds and other organic wastewater contaminants in a conventional drinking-water-treatment plant. *Science of the Total Environment*, **329**(1–3), 99–113.

Stedmon, C. A., & Bro, R. (2008). Characterizing dissolved organic matter fluorescence with parallel factor analysis: A tutorial. Limnology and Oceanography: *Methods*, **6**(11), 572–579.

Stedmon, C. A., & Cory, R. M. (2018). Biological Origins and Fate of Fluorescent Dissolved Organic Matter in Aquatic Environments. *Aquatic Organic Matter Fluorescence*, 278–300.

Stedmon, C. A., Markager, S., & Bro, R. (2003). Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry*, **82**(3–4), 239–254.

Steiner, E. C., Pelletier, R. R., & Trucks, R. O. (1964). A Study of the Polymerization of Propylene Oxide Catalyzed by Anhydrous Potassium Hydroxide. *Journal of the American Chemical Society*, **86**(21), 4678–4686.

Stuer-Lauridsen, F. (2005). Review of passive accumulation devices for monitoring organic micropollutants in the aquatic environment. *Environmental Pollution*. **136**(3), 503-524.

Sun, L., Perdue, E. M., Meyer, J. L., & Weis, J. (1997). Use of elemental composition to predict bioavailability of dissolved organic matter in a Georgia river. *Limnology and Oceanography*, **42**(4), 714–721.

Tautenhahn, R., Bottcher, C., & Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, **9**, 16.

Thermo Scientific. (2009). HESI-II Probe User Guide.

Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., Weyhenmeyer, G. A. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, **54**(6), 2298–2314.

Vasquez, M. I., Lambrianides, A., Schneider, M., Kümmerer, K., & Fatta-Kassinos, D. (2014). Environmental side effects of pharmaceutical cocktails: What we know and what we should know. *Journal of Hazardous Materials*, **279**, 169-189**.**

Verlicchi, P., Al Aukidy, M., Galletti, A., Petrovic, M., & Barceló, D. (2012). Hospital effluent: Investigation of the concentrations and distribution of pharmaceuticals and environmental risk assessment. *Science of the Total Environment*, **430**, 109–118.

Viant, M. R. (2008). Recent developments in environmental metabolomics. *Molecular BioSystems*, **4**(10), 980–986.

Viant, M. R., & Sommer, U. (2013). Mass spectrometry based environmental metabolomics: A primer and review. *Metabolomics*, **9**, 144–158.

Vrana, B., Allan, I. J., Greenwood, R., Mills, G. A., Dominiak, E., Svensson, K., Morrison, G. (2005). Passive sampling techniques for monitoring pollutants in water. *TrAC - Trends in Analytical Chemistry*. **24**(10), 845-868.

Wang, Y., Yang, G., Dai, D., Xu, Z., Cai, L., Wang, Q., & Yu, Y. (2017). Individual and mixture effects of five agricultural pesticides on zebrafish (Danio rerio) larvae. *Environmental Science and Pollution Research*, **24**(5), 4528–4536

Weeks, J. A., Guiney, P. D., & Nikiforovz, A. I. (2012). Assessment of the environmental fate and ecotoxicity of N,N-diethyl-m-toluamide (DEET). *Integrated Environmental Assessment and Management*, **8**(1), 120–134.

Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., & Mopper, K. (2003). Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environmental Science and Technology*, **37**(20), 4702–4708.

Wood, T. P., Du Preez, C., Steenkamp, A., Duvenage, C., & Rohwer, E. R. (2017). Database-driven screening of South African surface water and the targeted detection of pharmaceuticals using liquid chromatography - High resolution mass spectrometry. *Environmental Pollution*, **230**, 453–462.

Wünsch, U. J., Murphy, K. R., & Stedmon, C. A. (2015). Fluorescence Quantum Yields of Natural Organic Matter and Organic Compounds: Implications for the Fluorescence-based Interpretation of Organic Matter Composition. *Frontiers in Marine Science*, **2**, 98.

Yates, C. A., & Johnes, P. J. (2013). Nitrogen speciation and phosphorus fractionation dynamics in a lowland Chalk catchment. *Science of The Total Environment*, **444**, 466–479.

Yates, C. A., Johnes, P. J., & Spencer, R. G. M. (2016). Assessing the drivers of dissolved organic matter export from two contrasting lowland catchments, U.K. *Science of the Total Environment*, **569–570**, 1330–1340.

Ying, G. G. (2006). Fate, behavior and effects of surfactants and their degradation products in the environment. Environment International, **32**(3), 417–431.

Zabiegała, B., Kot-Wasik, A., Urbanowicz, M., & Namieśnik, J. (2010). Passive sampling as a tool for obtaining reliable analytical information in environmental quality monitoring. *Analytical and Bioanalytical Chemistry*, **396**(1), 273–296.

Zenker, A., Cicero, M. R., Prestinaci, F., Bottoni, P., & Carere, M. (2014). Bioaccumulation and biomagnification potential of pharmaceuticals with a focus to the aquatic environment. *Journal of Environmental Management*, **133**, 378–387.

Zgoła-Grześkowiak, A., Grześkowiak, T., Zembrzuska, J., & Łukaszewski, Z. (2006). Comparison of biodegradation of poly(ethylene glycol)s and poly(propylene glycol)s. *Chemosphere*, **64**(5), 803–809.

Zhang, F., Harir, M., Moritz, F., Zhang, J., Witting, M., Wu, Y., … Hertkorn, N. (2014). Molecular and structural characterization of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry. *Water Research*, **57**, 280–294.

Zhang, Z., Hibberd, A., & Zhou, J. L. (2008). Analysis of emerging contaminants in sewage effluent and river water: Comparison between spot and passive sampling. *Analytica Chimica Acta*, **607**(1), 37–44.

Zhou, J. L., Maskaoui, K., & Lufadeju, A. (2012). Optimization of antibiotic analysis in water by solid-phase extraction and high performance liquid chromatography-mass spectrometry/mass spectrometry. *Analytica Chimica Acta*, **731**, 32–39.

Zuccato, E., & Castiglioni, S. (2009). Illicit drugs in the environment. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **367**(1904), 3965–3978